

Monographic library “Knowledge and business”, book 19

Natalia Marinova

ARTIFICIAL GENERAL INTELLIGENCE SYSTEMS CHALLENGES

2023

Publishing house “Knowledge and business” Varna

This book or any part of it may not be copied or distributed electronically without the written permission of the author.

- © Natalia Marinova, author, 2023.
- © Publishing house "Knowledge and business", 2023.

This monograph is indexed in RePEc
(<https://econpapers.repec.org/bookchap/kabmonogr/19.htm>)

ISBN 978-619-210-068-1

Editorial board "Knowledge and business"

- Prof. PhD Petko Shterev Iliev – Head editor, University of Economics Varna, Bulgaria
Assoc. Prof. PhD Svetlozar Dimitrov Stefanov – Deputy Head editor, University of Economics Varna, Bulgaria
Prof. PhD Julian Andreev Vasilev – Deputy Head editor, University of Economics Varna, Bulgaria
Assoc. Prof. PhD Anastasia Stefanova Konduktorova – Scientific Secretary, University of Economics Varna, Bulgaria
Prof. PhD Marin Todorov Neshkov, University of Economics Varna, Bulgaria
Assoc. Prof. DrSc. Pavel Stoyanov Petrov, University of Economics Varna, Bulgaria
Assoc. Prof. PhD Sabka Dimitrova Pashova, University of Economics Varna, Bulgaria
Assoc. Prof. PhD Andriyana Andreeva, University of Economics Varna, Bulgaria
Assoc. Prof. PhD Desislava Borislavova Serafimova, University of Economics Varna, Bulgaria
Chief Assistant Prof. PhD Todor Kostadinov Dyankov, University of Economics Varna, Bulgaria
Chief Assistant Prof. PhD Svetlana Todorova, University of Economics Varna, Bulgaria
Prof. PhD Zdzislaw Polkowski, Uczelnia Jana Wyżykowskiego, Polkowice, Poland
Prof. PhD Stefan Bojnec, University of Primorska, Koper, Slovenia
Prof. PhD Young Moon, Syracuse University, Institute for Manufacturing Enterprises, USA
Prof. PhD Rajesh Khajuria, Gujarat Technological University, Ahmedabad, India
Dr. Amin Parag, SIES Colleague of Management Studies, Navi Mumbai, India
Assoc. Prof. Dr Eduard Stoica, University Lucian Blaga of Sibiu

ARTIFICIAL GENERAL INTELLIGENCE SYSTEMS CHALLENGES

Natalia Marinova¹

¹D. Tsenov Academy of Economics, Svishtov, Bulgaria

n_marinova@uni-svishtov.bg

Abstract

The **subject** of this study is researching the challenges of artificial general intelligence systems – systems that can independently solve problems from different domains of human life. The **purpose** of this review monographic research is to explore the nature, application and risks of current artificial narrow intelligence systems and the possibility of their evolution into solutions with general intelligence.

According to the purpose thus formulated, we direct our efforts towards the following **tasks**:

1. Analysing the development of the Artificial Intelligence field and describing the main research approaches in it.
2. Highlighting the capabilities and domains of artificial narrow intelligence systems.
3. Systematization of the methods, principles and algorithms implemented in solutions with narrow intelligence.
4. Conceptualising the 'general intelligence' characteristic and the challenges of systems with such feature.
5. Dividing hazards of artificial narrow intelligence systems into several key points.
6. Systematization of regulatory tools and effects guiding the development of ethical artificial intelligence systems.

The main research **thesis** of the paper is that despite the undeniable evolutionary development of artificial intelligence technology since the beginning of the twentieth century, the implementation of artificial general intelligence systems has not yet been proven possible and should be sought in a long-term time range.

Keywords: Artificial Intelligence, Artificial Narrow Intelligence systems, Artificial General Intelligence systems, Artificial Neural Networks, machine learning algorithms, deep learning, ethical machines, AI risks and challenges, AI applications

CONTENT

PREFACE	6
CHAPTER ONE. PREREQUISITES FOR THE EMERGENCE OF ARTIFICIAL GENERAL INTELLIGENCE	8
1.1. DEVELOPMENT OF THE INTERDISCIPLINARY SCIENTIFIC FIELD ARTIFICIAL INTELLIGENCE	8
1.2. RESEARCH APPROACHES IN THE FIELD OF ARTIFICIAL INTELLIGENCE	15
1.3. CAPABILITIES OF ARTIFICIAL NARROW INTELLIGENCE SYSTEMS	24
1.4. APPLICATION OF THE ARTIFICIAL NARROW INTELLIGENCE SYSTEMS	32
CHAPTER TWO. METHODS, PRINCIPLES, AND ALGORITHMS OF ARTIFICIAL NARROW INTELLIGENCE SYSTEMS	42
2.1. METHODS FOR SEARCHING, OPTIMISATION, AND CLASSIFICATION	42
2.2. PRINCIPLES FOR LOGICAL AND PROBABILISTIC REASONING.....	47
2.3. MACHINE LEARNING ALGORITHMS	50
2.4. ARTIFICIAL NEURAL NETWORKS	55
CHAPTER THREE. CHALLENGES AND RISKS OF ARTIFICIAL GENERAL INTELLIGENCE SYSTEMS.....	77
3.1. CONCEPTUALIZATION OF GENERAL INTELLIGENCE	77
3.2. CHALLENGES OF ARTIFICIAL GENERAL INTELLIGENCE SYSTEMS.....	85
3.3. RISKS OF ARTIFICIAL NARROW INTELLIGENCE SYSTEMS	91
3.4. ARTIFICIAL INTELLIGENCE ETHICS AND REGULATIONS.....	100
CONCLUSION.....	108
BIBLIOGRAPHY	110

Preface

People for millennia have dreamed of creating machines to replace them in the performance of various activities. Since the advent of the first electronic computing machines in the mid-1950s, several researchers have begun attempts to connect computer technology with human consciousness, laying the groundwork for the logical-symbolic modelling approach of human intelligence and forming a new field in Computer Science called Artificial Intelligence (AI).

Today's computing machines can generate not only data, but also knowledge. The formal layout of this knowledge as ideas, strategies and solutions to real problems expresses the main purpose of creating artificial intelligence systems – to perform no worse (and possibly better) than humans activities thought to require intelligence (Nilsson N. J., 2010, p. 646). Achieving this goal requires systems of this class to have the capabilities to correctly interpret data from their surroundings, to learn from this data and to flexibly apply what they learn when solving specific tasks.

Modern research in the field of Artificial Intelligence aims to solve several basic theoretical-applied tasks:

- ① Coding in computer programs the non-numerical and implicit aspects of solving a problem as it encodes its mathematical and algorithmic aspects.
- ② Creation of automated computer systems that in their behaviour resemble (mainly functionally and not so structurally, anatomically, and physiologically) human thought processes (for example, perception, decision making, induction, deduction, analogy).
- ③ Searching for new approaches in data and information processing and expanding the range of computer-assisted tasks with a focus on informal ones.
- ④ Adequate modelling of human reasoning based on the study of natural intelligence and the principles of brain functioning.

In recent decades, the artificial intelligence systems have developed quite intensively at a pace comparable in scale to the Internet revolution. This form of application of computing is an **actual** trend in modern scientific and practical research, since:

- A lot of scientific efforts are concentrated in this scientific strand.
- In the process of developing artificial intelligence systems, new methods for conducting interdisciplinary research are discovered and new views are formed on the role of certain scientific achievements.
- The goal of reproduction of various intellectual human tasks makes the field of Artificial Intelligence universal.

Currently, artificial intelligence systems are used in several domains of human life such as Medicine, Mathematics, Software Engineering, Biology, Architecture, Automotive, Logistics, Warfare, etc., where we daily witness new and revolutionary achievements. Although the trend in the evolutionary development of artificial intelligence solutions is upward, the technology is still narrow profiled and does not reach human capabilities.

This review monographic study is based on Kaplan and Haenlein's concept (Kaplan & Haenlein, 2019) which divides the progress in the field of Artificial Intelligence into three evolutionary stages:

1. Artificial Narrow Intelligence - artificial intelligence systems with the help of which problems of a certain domain of human life are solved.

2. Artificial General Intelligence - artificial intelligence systems that can independently solve problems from different domains of human life.

3. Artificial Super Intelligence - artificial intelligence systems that can be applied in any domain of human life, requiring problem solving through scientific creativity, social skills, and wise thinking.

The opinions of the researchers are divided on the question of whether the goals of general and super-intelligence should necessarily be pursued or whether as many specific problems as possible should be solved in the hope that narrow solutions will indirectly lead to the realization of the main goal of the Artificial Intelligence field. The characteristic 'general human intelligence' is difficult to define and measure, while the successes of modern artificial intelligence systems with narrow intelligence in solving specific problems can easily be measured by quantitative metrics.

The successes of applying artificial intelligence systems in the tasks of creating robotic vehicles, automatic scheduling and planning, machine translation, natural language recognition, recommendations, image understanding, medical diagnostics and dealing with climate problems are accompanied by several risks and challenges from the use of the technology. The close interrelationship of this scientific strand with Philosophy has given rise to discussions about the limits of machine possibilities, the opportunity of creating artificial consciousness and brain, the equivalence of human and machine intelligence, and the potential behaviour of intelligent machines. While the human species is currently dominant because of its brain's better cognitive abilities, many researchers worry that programming human virtues in machines will prove an impossible task.

The **purpose** of this review monographic research is to explore the nature, application and risks of current artificial narrow intelligence systems and the possibility of their evolution into solutions with general intelligence.

According to the purpose thus formulated, we direct our efforts towards the following **tasks**:

- Analysing the development of the Artificial Intelligence field and describing the main research approaches in it.
- Highlighting the capabilities and domains of artificial narrow intelligence systems.
- Systematization of the methods, principles and algorithms implemented in solutions with narrow intelligence.
- Conceptualising the 'general intelligence' characteristic and the challenges of systems with such feature.
- Dividing hazards of artificial narrow intelligence systems into several key points.
- Systematization of regulatory tools and effects guiding the development of ethical artificial intelligence systems.

The main research **thesis** of the paper is that despite the undeniable evolutionary development of artificial intelligence technology since the beginning of the twentieth century, the implementation of artificial general intelligence systems has not yet been proven possible and should be sought in a long-term time range.

Chapter One.

Prerequisites for the Emergence of Artificial General Intelligence

1.1. Development of the Interdisciplinary Scientific Field Artificial Intelligence

Since John McCarthy's introduction of the term in 1956 as "the science and technology of making intelligent machines" in literature, we have found numerous definitions of systems (machines, computers, software, or agents) with artificial intelligence. Based on the tabular systematization of the different definitions made by Russell and Norvig (Russell & Norvig, 2016, p. 2), we present the theoretical definitions of artificial intelligence systems in the following two-dimensional matrix (Figure 1.1):

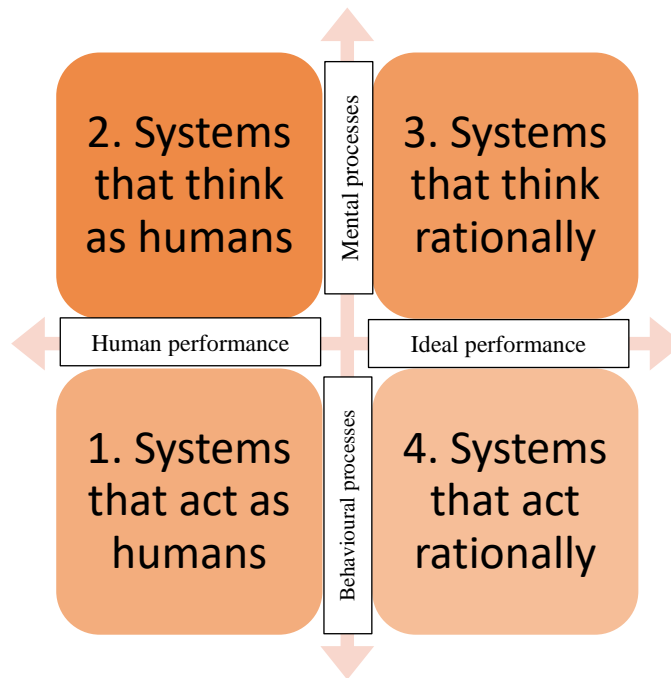


Figure 1.1. Categories of definitions of the term 'Artificial Intelligence'.

Source: author's illustration of Russell and Norvig's grouping.

The mental dimension of the matrix is focused on the capabilities of the artificial intelligence systems to mimic the biological processes of thinking and understanding, which can only be achieved by modelling the internal structure and the principles of data processing in an existing natural system through some technical means.

The behavioural dimension of the matrix focuses on the outcome of the functioning of the artificial intelligence systems, i.e., achieving a match in the behaviour of artificially created intelligent systems with naturally existing ones¹. The intelligent capabilities of natural systems are imitated entirely by means of computing.

¹One of the reasons for the presence in the scientific literature of many different definitions of the term 'Artificial Intelligence' is, in our opinion, the lack of a unified definition of the term 'intelligence' itself, which is seen as a skill in solving complex problems, as the ability to learn or as an opportunity to interact with the surrounding environment through communication, perception and awareness of the perceived. Since intelligence is primarily associated with natural abilities to think, create, and respond correctly in a given situation, it is obvious that these properties must be possessed to one degree or another by any artificial intelligence system.

The human-centric dimension of the matrix borrows postulates from the behavioural sciences by seeking to answer whether artificial intelligence should simulate natural intelligence, whether the study of human biology has the same importance for artificial intelligence systems as it has for example, the knowledge of bird biology for aeronautical engineering, etc.

The rational dimension of the matrix seeks answers to the questions whether the intelligence abilities of artificial intelligence systems can be reproduced only through symbols or whether numerical processing is also required, whether artificial intelligence systems should be able to solve a large number of completely unrelated problems, can intelligent behaviour be described using simple and well-known principles, etc.

Each of the four categories of definitions has its adherents and expresses the ideas and tools of some of the main approaches developed over the years in the study of artificial intelligence systems.

According to the first category of definitions, artificial intelligence systems can be considered acting as human beings only if they cover the well-known Turing test. This standard for assessing the behaviour of a computer system as indistinguishable from human life is still used to measure the degree of intelligence of artificial intelligence systems.

Proponents of the second category of definitions focus mainly on the presence in computing machines of the mental ability to think as the highest degree of intelligence in information processing. According to them, the potential of Thinking Machines to learn and change their behaviour in order to perform basic cognitive activities has made them the most complex and universal artificial intelligence systems invented by man.

Followers of the third category of definitions believe that rationality in the artificial intelligence systems "thinking" is achieved only by embedding certain structural models of arguments that always return precise conclusions under properly set conditions. Their ideas are primarily based on the application of principles from formal logic, heuristic search, and fuzzy logic.

The latter category of definitions focuses on the ability of artificial intelligence systems to act correctly on the basis of available data in order to achieve the best possible or expected (in conditions of uncertainty) result. Rational computer agents act on the basis of well-founded logical and mathematical formalisms.

As you can see, there is no established theory or paradigm in the field of Artificial Intelligence that unites the opinions of different scientists. Modern mathematical logic neither proves nor denies the possibility of creating artificial general intelligence systems. So far, there are no principal limitations to modelling natural intelligence, thinking and various intellectual functions on artificial technical systems. And according to some researchers (Haugeland, 1985) such an artificial imitation is not even necessary since machines possess synthetic intelligence - intelligence in its true form.

The considered categories of definitions reflect the interdisciplinary nature of the field of Artificial Intelligence, which over time has been influenced by the ideas, viewpoints, and techniques of different branches of scientific knowledge.

Philosophical schools of thought rationalism, dualism, materialism, and logical positivism pose fundamental questions about the use of formal logical rules, the formation of reason, the origin of human knowledge and its relationship to taking action, etc.

Mathematics formalizes the direction of Artificial Intelligence through the ideas of propositional and first-order logic, logical deductions, Theory of reference and Theory of probability.

Economics, as a science studying the way decisions leading to the maximization of desired outcomes, influences the field of Artificial Intelligence with its Utility theory, Decision theory, Game theory, Satisfaction theory, and Operations research.

The science of the biological nervous system Neurology assists in the search for the answer to the question of how (by analogy with computer systems) neurons and sensory centres in the brain process information. Attempts to compare the human brain and the digital computer by basic features (Table 1.1) prove that at present even a machine with virtually unlimited capacity cannot pass the threshold of singularity (the point after which computer possibilities will surpass human capabilities).

Table 1.1. Comparison of the capabilities of the human brain with a personal and supercomputer.

Source: Russel, S., & Norvig, P. (2021). Artificial Intelligence: A modern approach (4th ed.). Essex: Pearson Education Limited. p. 54.

Feature	Supercomputer	Personal computer	Human brain
Number of computational units	10 ⁴ computational units 10 ¹² transistors	4 computational units 10 ⁹ transistors	10 ¹¹ neurons
Number of storage units	10 ¹⁴ bytes RAM 10 ¹⁵ bytes disk	10 ¹¹ bytes RAM 10 ¹³ bytes disk	10 ¹¹ neurons 10 ¹⁴ synapses
Cycle time	10 ⁻⁹ seconds	10 ⁻⁹ seconds	10 ⁻³ seconds
Number of operations per second	10 ¹⁵	10 ¹⁰	10 ¹⁷
Number of memory refresh operations per second	10 ¹⁴	10 ¹⁰	10 ¹⁴

Psychology tries to explain what gives rise to the processes of thinking and acting in humans and animals. According to the ideas of Cognitive psychology, the brain is seen as an information processing device that translates the sensory stimulus into some form of internal representation, which is manipulated by cognitive processes to extract a new internal representation, and finally both are retranslated into action.

Computer engineering makes artificial intelligence systems complete, "equipping" them with automatable and programmable hardware. In the role of such most often enter different types of digital electronic computers.

The autonomous functioning of full-fledged artificial intelligence systems is provided by various biological and mechanical self-regulating and self-controlling mechanisms (Control theory), through which an artificial intelligent machine with stable adaptive behaviour can be created (Cybernetics).

Linguistics introduces the idea of the relationship between natural languages and thinking. The combination of concepts from the two scientific strands creates the hybrid scientific field Computational linguistics (natural language processing).

The need to know many postulates, techniques, mechanisms, and tools from the listed basic scientific fields of human knowledge predetermines the difficulty in the theoretical study and practical implementation of artificial intelligence systems. Some limitation of the conceptual apparatus used is found in the view of Stuart Shapiro, basing the research approaches in artificial intelligence on the following sub-strands (Figure 1.2):

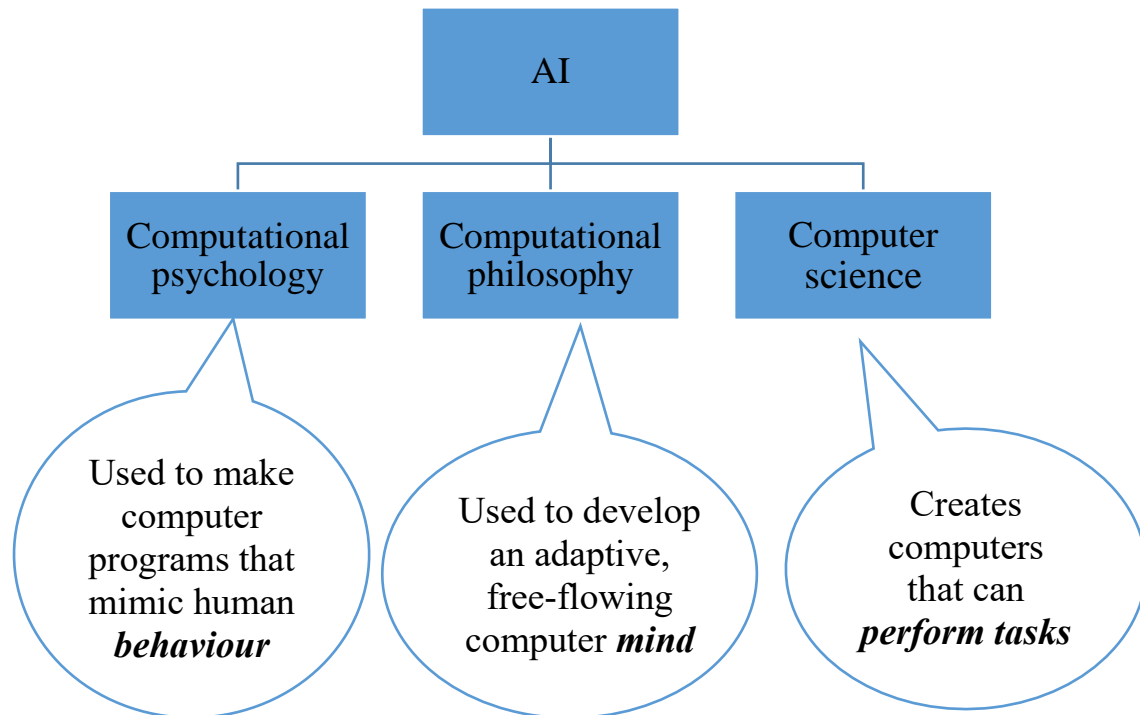


Figure 1.2. Scientific strands on which research approaches in Artificial Intelligence are based.

Source: *Artificial Intelligence / An Introduction*. (n.d.). Retrieved January 28, 2022, from <https://www.geeksforgeeks.org/artificial-intelligence-an-introduction/>

The versatility of studies and viewpoints in the field of artificial intelligence is inevitably due to the rich chronology in the evolutionary development of artificial intelligence systems, part of which we have illustrated below (Figure 1.3).

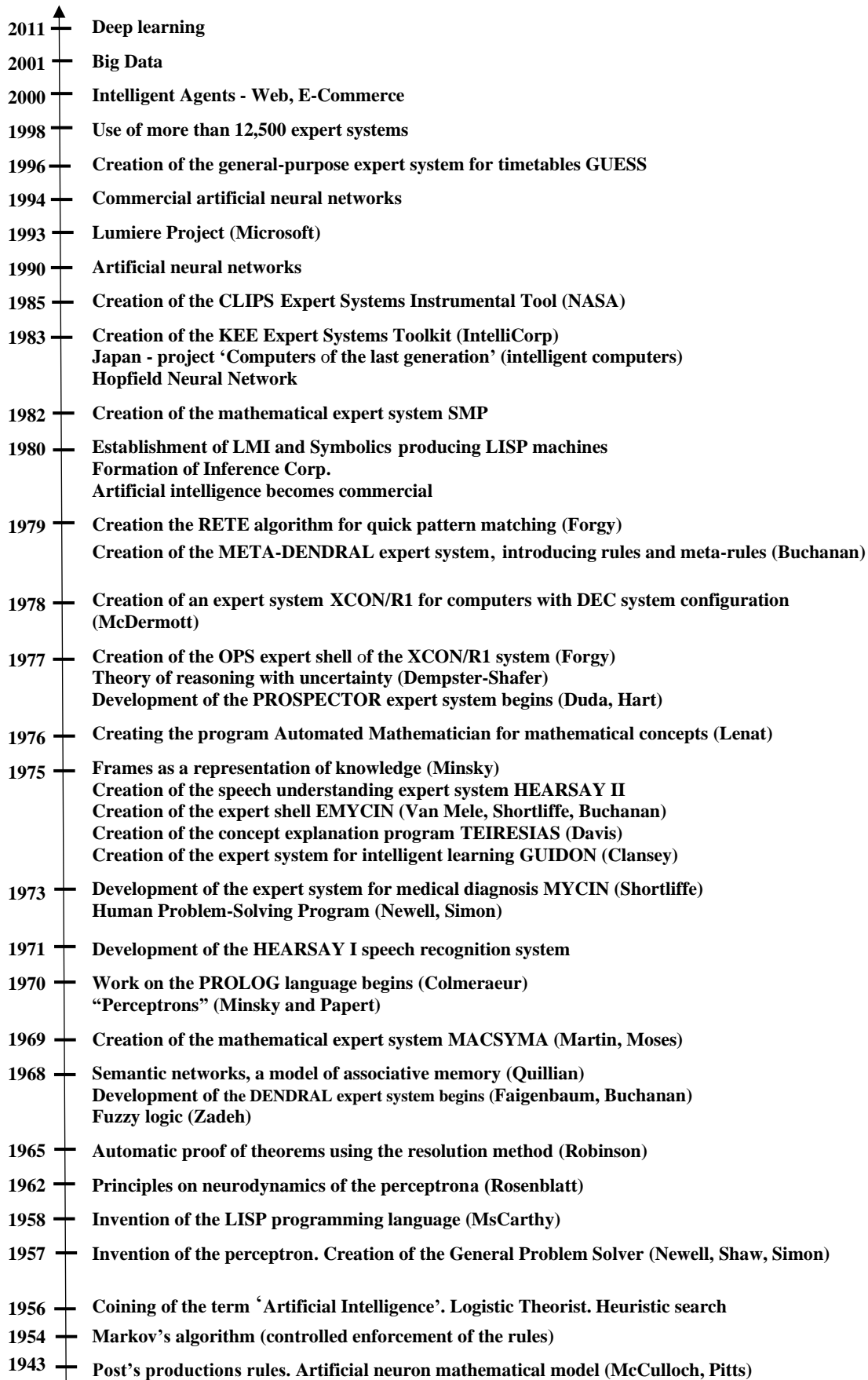


Figure 1.3. Key events in the evolution of artificial intelligence systems.

Source: author's illustration.

From the timeline listing key events in the development of artificial intelligence, it is clear that only in a period of 80 years remarkable progress has been made. Although we find desires and attempts to embody intelligent abilities in artificially created beings and automata in ancient times, in the ideas of ancient Greek philosophers, in the mechanical calculating machines of Renaissance scientists, and in the abstract Turing Machine², the beginning of modern artificial intelligence systems is considered to be the creation of the mathematical model of an artificial neuron³.

John McCarthy's introduction of the term 'Artificial Intelligence' during a two-week seminar at Dartmouth College in 1956 "institutionalized" artificial intelligence as a distinct branch of scientific knowledge. The demonstration of the Logistic Theorist (the first computer program with symbolic reasoning capabilities) and the discussions between participants with interests in the field of automata theory, artificial neural networks and the study of intelligence inspired future research on the subject and marked the beginning of a twenty-year period of artificial intelligence systems creation in the following directions:

1) Development of general-purpose systems based on symbolic reasoning - programs for algorithmic incremental solution of mathematical problems, games, and problems of everyday human life (General Problem Solver).

2) Development of knowledge-based systems for a particular area – a hypothetical self-learning Advice Taker program (McCarthy, 1958) using general knowledge in problem solving.

3) Using microworlds in solving problems requiring intelligence - the most popular microworld of Marvin Minsky and Seymour Papert is composed of coloured blocks⁴ of different shapes and sizes, placed on a flat surface, which can be displaced in a certain way by a robotic arm.

4) Implementation in computers of the possibility of communicating in natural language - semantic networks, ELIZA (the first program with artificial intelligence communicating in English).

5) Artificial neural networks – perceptron, an artificial neural network ADALINE for pattern recognition.

6) Robotics – unimate industrial robot, Shakey the Robot with moving, perception and problem-solving capabilities, an WABOT-1 intelligent humanoid robot.

Despite the implementation of many successful artificial intelligence systems, underestimating the complexity of the problems solved and not materializing the promised results criticizes and stops funding a large part of the researchers for a certain period. A significant factor in this decline in interest in the Artificial Intelligence strand is the

² The famous article "Computing machinery and intelligence" (Turing, 1950) explains the concepts of 'Turing test', 'machine learning', 'genetic algorithms' and 'reinforcement learning', inspires the idea of creating an electronic brain and underpins the future Artificial Intelligence strand.

³ The ideas of an artificial neuron and a network of logically connected neurons in the article "A logical calculus of the ideas immanent in nervous activity" by Warren McCulloch and Walter Pitts were later further developed by Donald Hebb, Marvin Minsky and Dean Edmonds (McCulloch & Pitts, 1943).

⁴ The subsequent studies of Gerald Sussman, Adolfo Guzman, David Waltz, and Patrick Winston in the field of computer vision are based on the block paradigm.

compromise of perceptrons as common computing components⁵.

Ideas for overcoming the shortcomings of incremental algorithms and the problem of combinatorial explosion scientists found in knowledge as a foundation for creating computing machines with reasoning capabilities in a given domain. The need to implement different types of knowledge in artificial intelligence systems prompted the creation of more formalisms⁶ for their presentation (horn clauses, frames, and scenarios), the development of a number of expert systems (DENDRAL, MYCIN, META-DENDRAL, EMYCIN and tens of thousands of others) and several national and international initiatives for the creation of new generation computers. The extraordinary success of the first commercial expert system R1 marks the beginning of the Artificial Intelligence industry.

The creation of the two-layer artificial neural networks and the artificial neural networks back-propagation training algorithm attracts again the research interest in the Connectionism strand. The widespread application of the new technology in optical character and speech recognition tasks commercialized it and led to a decline in the use of expert systems.

At the same time, the idea of an entirely new approach to artificial intelligence based on Robotics is beginning to be proclaimed⁷. The empirical proof of several ideas adopted by similar scientific fields (Information theory, Stochastic modelling, Mathematical optimization, Control theory, Statistical analysis, etc.) finally turns the field into a formal scientific branch of human knowledge.

Before the beginning of the new millennium, the artificial intelligence systems achieved some of the goals enthusiastic founders of the division – programs that beat humans in various games, competitions between robots, the sale of entertainment robotic devices, etc. The emergence of the intelligent agent's paradigm makes it possible for artificial intelligence systems to enter the web (through search engines, directories, and recommendation systems) and successfully implement them in industries such as Logistics, Industrial Robotics, Medicine, Data Mining, etc.

The evolution of artificial intelligence systems since 2000 has been influenced by the widespread use of the Global Web in all spheres of public life, the availability of many sets of Big Data, affordable and fast computing systems, and advances in deep machine learning techniques. Due to the impossibility of exhaustively listing all achievements of the strand, in Figure 1.4 below we have selected only basic ones:

⁵ The book "Perceptrons" by Marvin Minsky and Seymour Papert from 1969 proves their inability to recognize the logical function XOR and literally stops research in the entire Connectionism strand (an approach in Cognitive Science that seeks to explain the thinking process using artificial neural networks) for a period of 10 years.

⁶ Knowledge representation models represent a set of syntactic and semantic conventions that make it possible to describe objects, properties, processes, etc. in a particular domain (Маринова, 2014, стр. 122).

⁷ According to Rodney Brooks and other researchers, a machine can only show real intelligence if it has a body with sensory-motor skills.

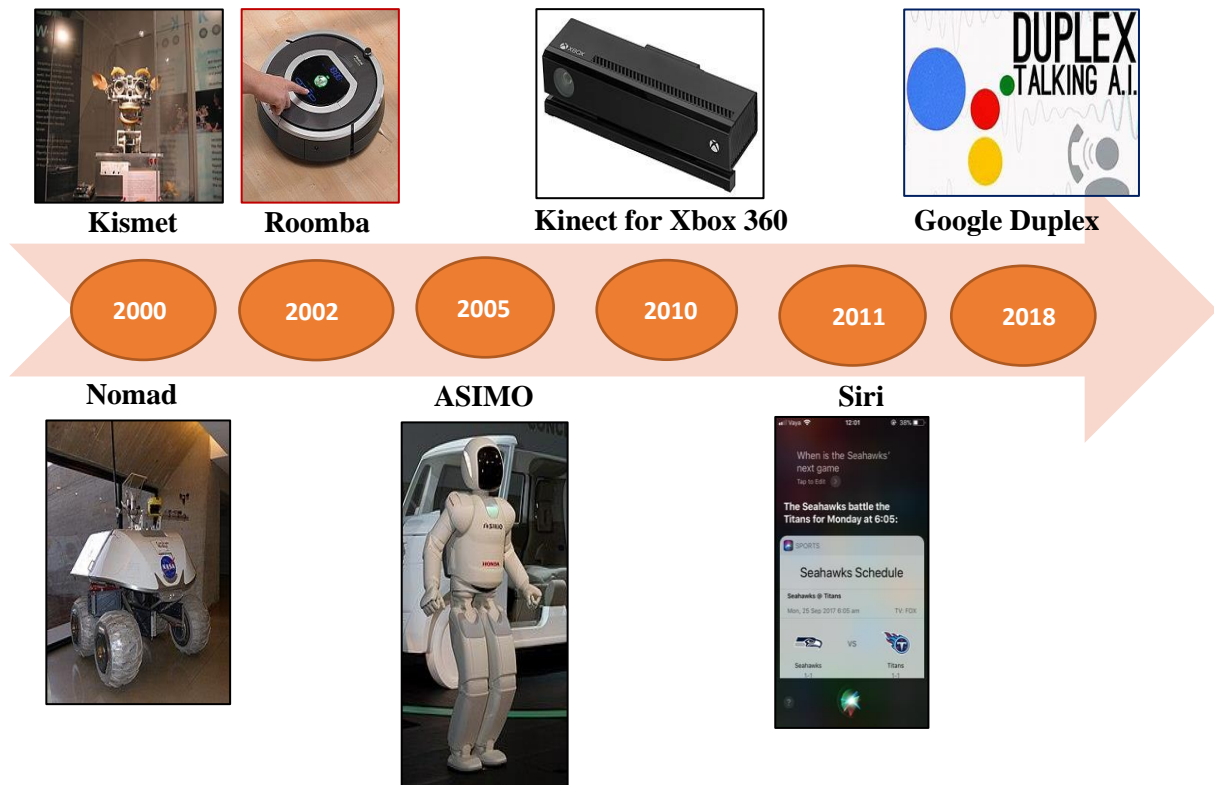


Figure 1.4. Key hardware and software artificial intelligence systems that have been developed since 2000.

Source: author's illustration.

1.2. Research Approaches in the Field of Artificial Intelligence

Tracking the evolutionary development of artificial intelligence systems over the years, we find several basic approaches on which research in the field is based – cybernetics, neuroscience, symbolic, computational, statistical, and intelligent agents’ approach. Undoubtedly, the successful implementation of artificial general intelligence systems is based on a common, and in some cases in-depth, knowledge of the ideas, techniques, and tools of each approach.

Cybernetics Approach

The formalization of the Artificial Intelligence strand as an independent scientific branch is preceded by the creation of simple cybernetic devices⁸ and attempts to overcome the limitations in the processing and transmission of encrypted signals through noisy channels (Shannon & Elwood, 1948).

Cybernetics establishes the unity of laws of government in living and non-living nature and introduces the idea that the neural structures of organisms’ function in a manner analogous to binary digital computers. Established in 1948 by Norbert Wiener, this management science studies the concepts of ‘learning’, ‘cognition’, ‘appearance’, ‘adaptation’, ‘social control’, ‘convergence’, ‘communication’, ‘effectiveness’, ‘efficiency’ and ‘connectivity’ in their abstract form rather than in the context of specific organisms or devices (Wiener, 1948).

⁸ William G. Walter's educational turtle robots, John Hopkin's mobile automated beast and Rosenblatt's perceptron are based on electronic circuits and exhibiting a simple form of intelligence.

According to cybernetics, the processes of storing, transmitting, and processing information in living systems and machines are similar (Figure 1.5). The series of physiological actions in biological organisms presented in the illustration closely resembles the functioning of a digital computer: as sensors in the natural system serve receptors, as a processing and coordinating movement component – the brain, and the role of an actuator enters the motor system of the body. Information in both types of systems is transmitted by electrical impulses, with only 10^{-16} joules consumed by natural energy versus 10^{-6} in computers.

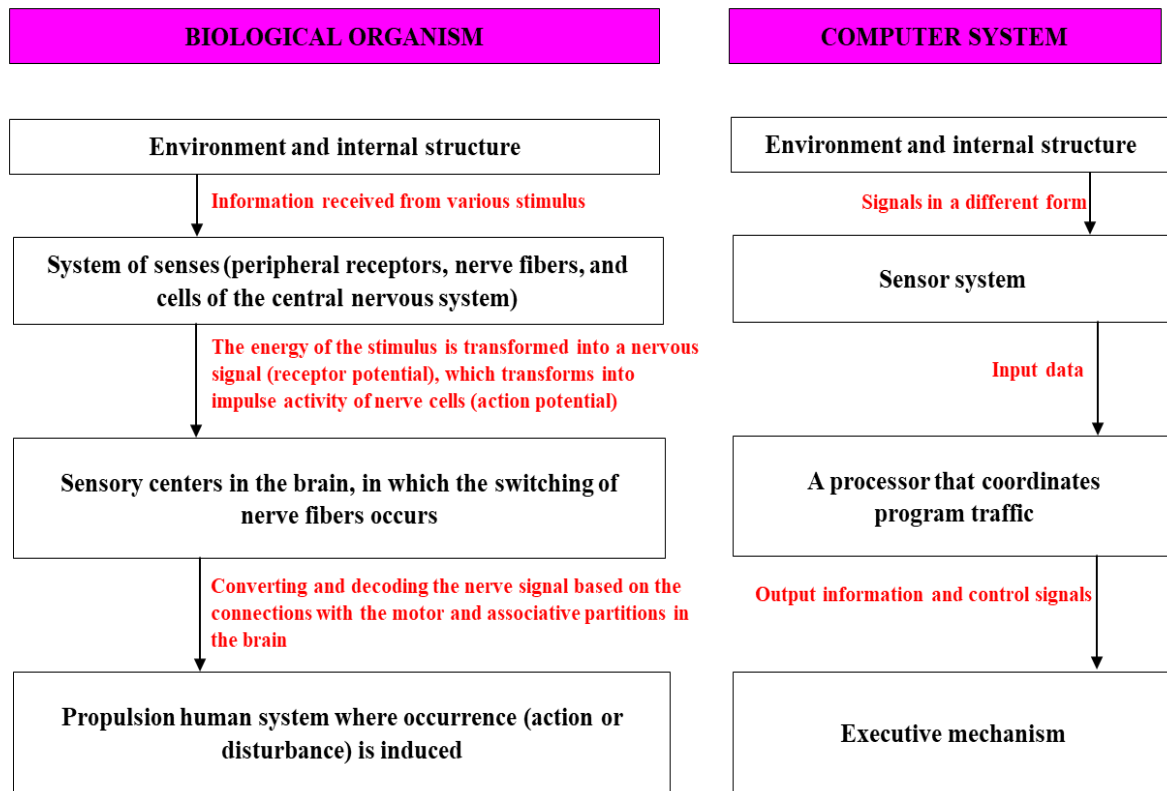


Figure 1.5. Actions performed in the sensory system of biological organisms and in binary digital computers.

Source: author's illustration of the comparison of cybernetics.

Neuroscience Approach

Important for research in the field of Artificial Intelligence is the neural doctrine, which considers the biological nervous system as a set of a huge number⁹ of sensory (afferent), motor (efferent) or connecting (inter-) neural cells. In the functioning of the individual parts of activated neural cells we again find an analogy with the computational processes conducted in computer systems (Figure 1.6). The ability to connect individual neurons into time-resistant structures (group of neurons, neural network, nervous system, mental activity, or brain) allows them to self-organize, to learn, and to manage the human body.

⁹ There are 100 billion neurons in the human brain alone, each connected to an average of up to 100,000 others. The number of inter neuronal connections is comparable to that of objects in the known universe.

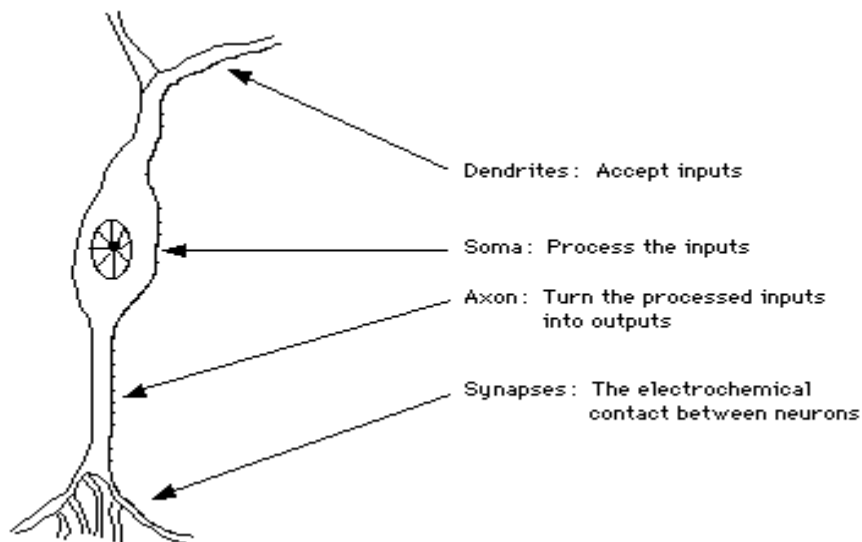


Figure 1.6. General structure of a typical neural cell.

Source: author's adaptation of a simplified illustration of the structure of a neural cell.

In individual parts or cells of the nervous system, no specific cognitive abilities for memorization, thinking or self-awareness have been found, suggesting that these are due to the overall organization and functioning of the various neural structures. A better understanding of the principles and mechanisms of operation of neural structures of natural systems needs to create devices with artificially emulated hardware-software topology. Philosophical and theoretical opinions on the possibility of creating artificial neural structures base these decisions either on traditional computer algorithms for storing, searching, and processing information, or on the need to create computer systems with a fundamentally new principle of operation.

Molecular Biology, Electrophysiology, Computational neurology¹⁰, and Connectionism are essential to advances in the study of composition, functioning, development, and change in the biological nervous system. According to connectionists (McClelland, Rumelhart, & Group, 1986) cognitive biological processes are due to simultaneously occurring distributed signalling activity in the neural connections of the brain, which can be digitally expressed and modified by some artificial neural network model¹¹. Networks of artificial neurons are the embodiment of the ideology of soft computing that attempt to imitate human reason through techniques of tolerance for inaccuracy and uncertainty, partial truth, and approximation (Zadeh, 1994).

The potential possibilities for structural-functional emulation of the vast and not yet sufficiently studied cognitive capacity of the biological brain on man-made artifacts to this

¹⁰ A branch of Neuroscience that uses mathematical models, theoretical analysis, and brain abstractions to describe the principles that govern the development, structure, physiology, and cognitive abilities of the nervous system (Schwartz, 1990). Computational neuroscience focuses on describing biologically realistic neurons and systems, rather than artificial ones.

¹¹ The existing variety of artificial neural network models obeys two basic principles: ① Each mental state can be described by a spatial vector of numerical values activating the neural units in the network. ② Memorization is done by modifying the strength of the connections (a scalar quantity called 'weight', represented in the form of a matrix) between the neural units, and the information in them is stored in the form of models for solving human problems.

day provoke attempts to create a new generation of computers¹², networks, associative storage devices and methods for parallel processing of information. Experimental attempts to develop a functioning artificial brain structure are based primarily on the use of artificial neural networks implemented on high-performance computing machines¹³. Successful emulation of a complete brain structure would undoubtedly be a key prerequisite for the invention of artificial general intelligence systems.

Symbolic Approach

The advent of digital computers shifts the focus of researchers in the field of Artificial Intelligence towards symbolic modelling¹⁴ of formal knowledge from different areas of human life. Symbolic intelligent systems apply new approaches to finding solutions, carry out symbolic reflections on knowledge and can be trained and adapted to newly created information (Atanasova, 2005, p. 33).

Summarizing the different research ideas based on symbolic representation of problems, concepts of logic and methods for searching for a solution, we can divide the studies in the field of symbolic artificial intelligence in the following directions:

1. Cognitive simulations.

Psychological experiments to study and formalize the human way of solving problems by Herbert Simon and Alan Newell (Simon & Newell, 1972) led to the development of a program for simulating the cognitive abilities of biological systems (General Problem Solver) and to the creation of the theory of physical symbolic systems. Their hypothesis represents symbols as related characters, and symbolic structures in the form of symbol lists, which can be manipulated by the computing machine through the means of special programming languages for processing lists (mainly LISP).

2. Logical approach.

Contrary to the first direction is the opinion of John McCarthy, who believed that machines should not simulate the human way of thinking but try to independently discover the essence of abstract reasoning and problem solving by applying algorithms based on formal logic. The development of this approach led to the creation of the programming language Prolog and the logical programming paradigm.

3. Anti-logical approach.

Proponents of the 'scruffy'¹⁵ approach believe that there is no single common principle that can cover all aspects of intelligent behaviour, and solving difficult tasks in the field of Artificial Intelligence requires specialized solutions. Examples of successful 'scruffy'

¹² Neuromorphic computers do not function on the basis of previously developed algorithms, but on the basis of specially selected examples for which they are trained.

¹³ The Swiss Blue Brain Project, for example, was able to simulate a single neocortical column in the brain of a rat by connecting 30 million artificial synapses.

¹⁴ The symbolic paradigm is designated by Haugeland (Haugeland, 1985) as a representative of the classic Artificial Intelligence with the abbreviation GOF AI – 'Good Old-Fashioned Artificial Intelligence'. The analogous term in Robotics is GOF R – 'Good Old-Fashioned Robotics'.

¹⁵ The designation of artificial intelligence research approaches as 'neat' and 'scruffy' was made by Roger Shank in 1973 (Shank, 1973). According to scientists working in the first direction, artificial intelligence solutions should be elegant, clear, and properly proven. The group of followers of the anti-logic approach believe that intelligence is too complex and computationally unsolvable to be achieved through any homogeneous system based on Logic.

solutions are the manually created knowledge base Cyc (Lenat & Guha, 1990) and the natural language processing programs ELIZA (Weizenbaum, 1976) and SHRDLU (Winograd, 1972).

The creation of computer systems with large integrated circuits facilitated the embedding of knowledge in artificial intelligence systems and the subsequent development and implementation of the first truly successful form of artificial intelligence software – the expert systems. Research efforts to embed knowledge in modern AI applications with narrow intelligence should be developed towards the implementation of subconscious (reflex, reactive) abilities of human behaviour.

Computational (Sub-symbolic) Approach

The computational approach in the field of Artificial Intelligence expands the capabilities of symbolic intelligent systems in the direction of imitating the processes of perception, movement, and manipulation of objects, learning and pattern recognition (Nilsson, 1998, p. 7). Attempts to emulate these natural abilities in the form of machine capabilities have created and developed several major modern scientific strands.

Artificial intelligence systems acquire and "understand" multidimensional digital and symbolic information from the real world through machine perception devices. The application of computer vision technology in the field of Artificial Intelligence aims at designing computer systems with learning capabilities in the recognition and interpretation of electromagnetic radiation in captured digital images or video (Ballard & Brown, 1982). Currently, computer vision systems are widely applied in the tasks of facial recognition, geographical modelling, scene reconstruction, event detection, video tracking, object recognition, visual servoing, tridimensionality pose estimation, learning, indexing, motion estimation, image restoration and aesthetic assessment¹⁶.

Artificial intelligence systems can move around and manipulate objects by interacting, embedding themselves in, and learning to respond to changes in their surroundings. According to scientists in the field of Robotics¹⁷, achieving such intelligence in machines requires the embodiment of bodily aspects¹⁸, deployment in a real or virtual environment and the availability of basic capabilities for modelling their own behaviour. Some¹⁹ even believe that machine intelligence can be generated organically without the need for programming with symbolic knowledge, but only because of the robot's simple interactions with the surrounding world.

The key research objectives of the Robotics field are:

¹⁶ Machine perception of data received from different sensors or entered by conventional computing means allows artificial intelligence systems to calculate the 'depth' metric to focus on a specific part of the image and distinguish variations in their illumination.

¹⁷ Robotics is simultaneously an applied science, borrowing knowledge from Computer Science, Cognitive Science, Control Theory, Electronics and Mechatronics, R&D activity, and technology industry.

¹⁸ According to the embodied mind thesis, many characteristics of cognition in biological systems (mental constructions of the highest level and the performance of various cognitive tasks) are due to the bodily aspects of organisms (motor system, sensory system, tactile interactions with the environment and innate instincts). Embodied agents interact with the environment through a physical or virtual body (avatar).

¹⁹ Rodney Brooks denotes this idea with the term 'Nouvelle AI'.

① Creation of new types of robots with innovative design²⁰ and/or domain (assembling, welding, heavy duty, etc.).

② Development of new methods to produce robotic devices.

③ Improving the dexterity and reliability of robotic devices in handling and interacting with fragile objects and living organisms. The possibilities of full-fledged proprioception inherent in biological systems that have evolved over millions of years still cannot be implemented in robots²¹. According to Hans Moravec, who noticed this paradox²² twenty-five years ago, robots with equivalent intelligent abilities will not be produced until 2040.

Modern theoretical and applied research in the field of Robotics, influenced by the open design²³ movement, evolutionary computations²⁴ methodology and software simulation²⁵ techniques, is developing in several directions:

① Construction of robots with different practical applications for industrial, commercial, educational (for training in engineering-focused educational specialties) or military use.

② Design of computer systems for sensor control and processing of the information collected by robots.

③ Creation of single prosthetic devices and complex exoskeletal structures.

④ Building smart environments (houses, cars, cities) equipped with different sensors and actuators.

⑤ Control of robotic swarms and of collectively and co-ordinately operating multi-robot systems.

The study of algorithms and statistical models through which computer systems solve specific problems relying only on templates and rules for inference is a basic concept in the field of Artificial Intelligence (Samuel, 1959). Artificial intelligence systems improve their performance based on implemented machine learning algorithms that work on the presumption that strategies, hypotheses, and inferences (logical, fuzzy, or probabilistic²⁶) that worked well in the past are likely to continue to work well in the future. They don't just learn

²⁰ The shape of many modern robots is inspired by the nature and characteristics of various species. Some models are purposefully constructed to resemble the human appearance to be more easily perceived when performing basic activities such as walking, lifting objects, speaking, learning, etc.

²¹ Perception of one's own movements and position (New sensors make for soft and sensitive robotic fingers, 2020).

²² In an interview (Moravec, 1997), he said that "it is difficult or even impossible for robots to give the perception and mobility skills of even one-year-old children, while it is relatively easy to make computers that can solve intelligence tests like older people or play checkers" and predicts the development of robot technology in four generations.

²³ Open-source robotics promotes the possibility of homemade and modification of robotic devices using publicly available hardware, blueprints, and schematics with free source code.

²⁴ Evolutionary robotics aims to improve the movement and behavior of robotic devices with a biological form by unprogrammed imitation of the process of natural evolution: the large population of robots competing in the performance of a task is gradually reduced and replaced until a unit with satisfactory behavior appears.

²⁵ Simulation robotics is used to improve the actions of real robots through learning in a virtual environment. Such software simulators save time and costs and provide the ability to directly implement on a finite physical device.

²⁶ According to the Ockham razor principle (the simplest theory that explains the data is the most likely one) the algorithm should be designed to prefer simpler theories, except in cases where the complex theory is proven significantly better than the simple one.

from data²⁷, they can also improve themselves by adopting new strategies or inference rules or writing other algorithms themselves.

From a theoretical point of view, the presence of data, time and memory with unlimited capacity would allow some of the machine learning algorithms to train themselves to approximate any mathematical function, to "remember" which combination of such would best describe the real world, to extract all possible knowledge to solve a problem, to deduce every possible hypothesis about the problem and to correlate it with the data. However, consideration of all existing hypotheses is impossible, which is why much of the research in the field of machine learning aims to identify and avoid hypotheses that are unlikely to happen or be useful.

Artificial intelligence systems can detect and recognize patterns in data and classify them into different categories (Bishop, 2006, p. 1). Pattern recognition systems are trained on sets of labelled and non-labelled data and use algorithms to search for the most likely output value corresponding to all statistical variations in input data (instances). Pattern recognition algorithms are classified according to the type of output data characteristics, the type of training (supervised or not) and its character (statistical or not). Many of these algorithms are probabilistic because they use statistical inference to find the best label for a particular instance.

Statistical Approach

Many of the classic artificial intelligence systems are based on the application of symbolic computing techniques (special algorithms and software for manipulating mathematical expressions and other mathematical objects), which, however, cannot model data from empirical research. This is the reason for the introduction of tools and methods from Statistics through which different types of cognitive architectures can be compared and unified and intelligent behaviour can be simulated²⁸.

The mathematical foundation of statistical means makes it possible to generate results from real data sets with a higher level of accuracy. Despite the demonstrated precise degree of measurement and reproduction of the results of empirical experiments, some researchers (Langley, 2011; Katz, 2012) consider that statistical formalization reduces the comprehensibility of results and does not support the long-term goal of creating artificial general intelligence systems.

²⁷ The ability of machines to train to perform a specific task based on data or experimental observation is studied by the Computational intelligence strand. At its core, it is an approach to dealing with complex real-world problems that mathematical or traditional modeling cannot solve for several reasons: processes are too complex for mathematical reasoning, the course of the process depends on unknown factors, the process may have a stochastic character (Siddique & Adeli, 2013).

²⁸ The term 'cognitive architecture' describes the theory of the structure of the human brain and its computational recreation into a comprehensible, formalized, and programmable computer model (Cohen, 1995).

Intelligent Agents' Approach

The main goal of the Artificial Intelligence strand is to create intelligently functioning computers and machines²⁹. Intelligent agents³⁰ are devices that perceive and analyse their surroundings and take actions that maximize the chance of successfully achieving their goals. The emergence of this paradigm integrates the other four main approaches in the field of Artificial Intelligence and becomes key to the theoretical-practical design and development of artificial narrow intelligence systems. The intelligent agents approach gives researchers a means of solving problems from specific application areas and of approbating the concept of abstract agents³¹.

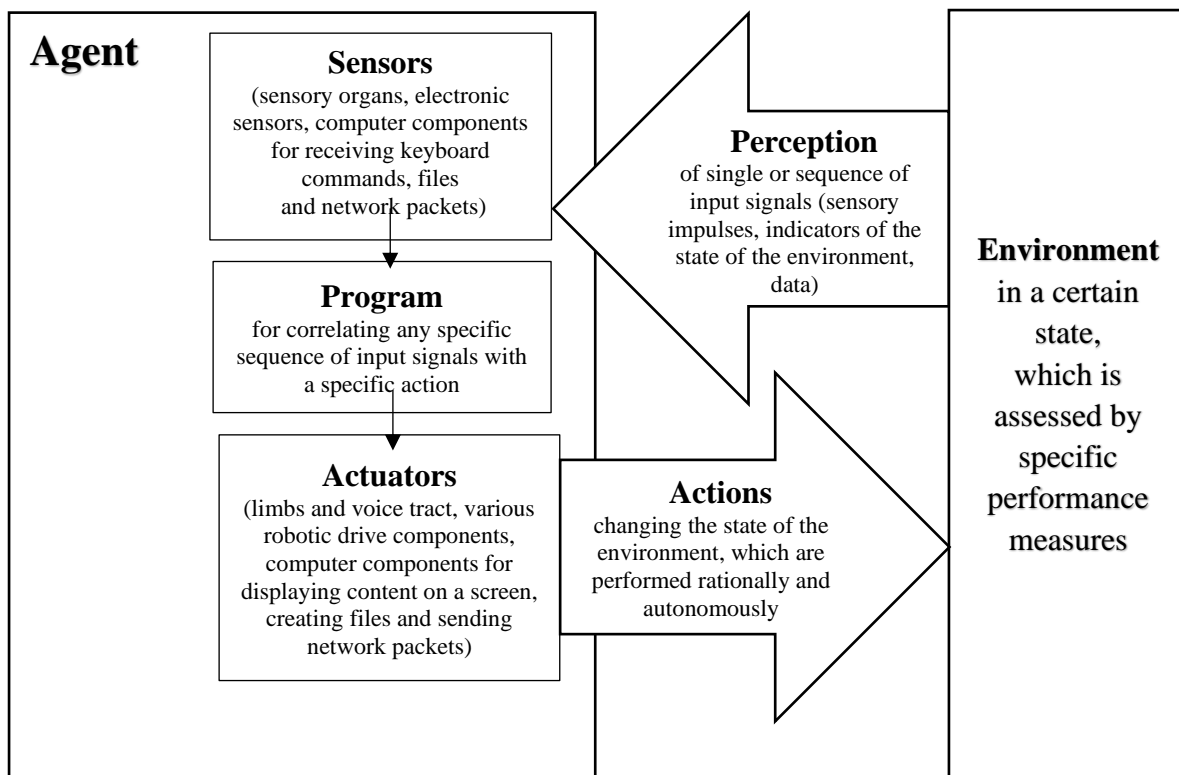


Figure 1.7. Intelligent agents' functioning.

Source: author's illustration.

Intelligent agents function in a specific state-changing environment (fully observable or partially observable, single-agent or multiagent, deterministic or stochastic, episodic or sequential, static or dynamic, discrete or continuous, known or unknown) that they perceive through sensors and with which they interact through activators (Figure 1.7). As a result of all signals registered at a given time, the execution of one or more actions is generated. The

²⁹ One of the main definitions of the Artificial Intelligence strand defines it as „science, which studies the synthesis and analysis of computational agents that act intelligently“ (Poole & Mackworth, 2017, p. 3).

³⁰ Unlike programs that simply perform tasks, computer agents operate autonomously, perceive their surroundings, survive for a long period of time, adapt to change, and create and pursue goals.

³¹ In Decision theory and Economics, intelligent agents are schematically described as an abstract functional system like a computer program, which is why they are sometimes referred to as abstract intelligent agents to distinguish them from their actual implementation as computer systems, biological systems, or organizations.

resulting actions must coincide with the desired ones and maximize the measures of environmental performance in order for the agent to be considered to be properly and rationally functioning.

The functioning of intelligent agents is conditioned by the following features:

① Inability to assess the degree of impact on environmental performance measures through a single universal indicator measuring the actions of agents to solve different tasks.

② Unknowing of all possible results of the implemented actions and inability to design agents with a pre-implemented execution of the most correct action.

③ Ability to train the agent to perform actions in new unknown environments based on pre-configured and acquired knowledge (any possible sequence of input signals at a given time) for familiar environments.

④ Autonomy in operation that does not depend on the agent's prior knowledge of the environment, which may be incomplete or inaccurate.

⑤ Designing to solve specific tasks. The creation of an intelligent agent to carry out robotic taxi transport, for example, requires consideration of the following characteristics (Table 1.2.):

Table 1.2. Characteristics of robotaxi.

Source: Russel, S., & Norvig, P. (2016). *Artificial Intelligence: A modern approach (3rd ed.)*. Essex: Pearson Education Inc. p. 40.

Performance measures	Environment	Actuators	Sensors
<ul style="list-style-type: none"> • Fuel costs and vehicle depreciation • Duration and cost of the trip • Violations of traffic rules and disturbance of other drivers • Safety and comfort of passengers • Profit 	<ul style="list-style-type: none"> • Types of roads • Elements of the road (cars, pedestrians, animals, repairs, police cars, puddles, and potholes) • Potential and current passengers • Climatic conditions • Local specifics of driving 	<ul style="list-style-type: none"> • Control of the engine via the accelerator, steering wheel, and brakes • Screen or voice synthesizer for communication with passengers • Means of communication with other vehicles 	<ul style="list-style-type: none"> • Controllable video surveillance cameras • Infrared or sonar sensors for measuring distance • Speedometer, odometer, and accelerometer • Sensors for monitoring the engine, fuel and electrical system • GPS system • Keyboard or microphone to set the destination

The behaviour of agents is described by a particular mathematical function. The systematization of all possible sequences of input signals and the response actions of the agent for each of them is algorithmized in the form of a program³². The components of the program represent the agent's application domain in an atomic (areas with unknown internal structure), factored (areas consisting of a vector of attribute values), or structured (areas containing

³² According to the type of program, five types of agents with different degrees of intelligence are distinguished: simple reflex agents, model-based reflex agents, goal-based agents, utility-based agents and learning agents.

entities with their own attributes and relationships to other entities) way. The implementation of programs requires the presence of a certain physical architecture, most often cognitive³³.

Recent developments in the field of intelligent agents are aimed at linking multiple interacting passive (without goal), active (with simple goals) or cognitive (performing complex calculations) agents in multiagent systems and at the creation of hybrid intelligent systems (neuro-symbolic systems, neuro-fuzzy systems, hybrid connectionist-symbolic models, fuzzy expert systems, connectionist expert systems, evolutionary neural networks, genetic fuzzy systems, etc.).

1.3. Capabilities of Artificial Narrow Intelligence Systems

As a general-purpose technology, the artificial intelligence systems are expected to be a multifaceted application in solving socially useful tasks that "promote democratic values such as freedom, equality and transparency" (Littman, et al., 2021). While there is no consensus on the domains best suited to artificial intelligence systems, intelligent agents are most often created that emulate human reasoning and problem-solving abilities, representing knowledge, automatically planning and creating schedules, learning, natural language understanding and processing, perceiving, moving and manipulating objects, expressing emotions, and creating works of art.

The human way of reasoning is based on common sense knowledge of basic physical categories (space, time, interactions) and on reflex mechanisms for explaining situations of reality, the behaviour and mental state of other people, the meaning of lexical categories in the natural language they use, etc. The implementation of such default reasoning in the artificial intelligence systems allows them to combine and rearrange the information in their knowledge base, to achieve a multitude of different results, to adopt new tasks in the form of explicitly described goals, to introduce and obtain new knowledge about their surroundings and to adapt to changes in the environment. Knowledge-based agents use principles from logic and language for declaratively (descriptively), procedurally (programmatically) or combined declarative and procedural representing, generating, and updating of signals received from the environment.

Intelligent problem solving through a random or fixed sequence of actions in artificial intelligence systems is based on the formulation of goals and on the search for the right sequence of actions to achieve them. Problem-solving intelligent agents are designed with a common search algorithm with capabilities to handle not only logically correct but also uncertain or incomplete information. Modern artificial intelligence systems solve problems by applying explicit or implicit symbolic reasoning, Bayesian inferences, analogizers, neural networks, or a combination of the four approaches together with other algorithms with or without artificial intelligence.

The full reasoning and problem-solving capabilities of artificial intelligence systems

³³ For each digital architecture with k number of bits for storing the program there exist exactly 2^k programs that need to be listed and tested in order to find the most correct one. While the cognitive capabilities of current computer architectures are not yet comparable to human intelligence, architectures such as the 4D-RCS Reference Model Architecture (Albus, 1993), ACT-R (Lebiere & Anderson, 1993), Soar (Laird, 2012), and LIDA (Franklin & Patterson, 2006) are seen as potential solutions for emulation of general intelligence in artificial intelligence systems.

depend on the comprehensiveness of their knowledge of the surrounding world (objects, properties, relationships between objects, processes, situations, events, mental states, time intervals, causes, effects, metaknowledge and other abstract categories of poorly researched knowledge domains). The implementation of the human way of handling common sense facts in intelligent agents through the techniques of ontological engineering is most often achieved by means of semantic networks and descriptive logics.

The development of correct and complete ontology enables artificial intelligence systems to index and return contextual content, to interpret situations, to make medical diagnoses, to find knowledge in large databases, to display automatic annotations, etc. The successful achievement of these tasks, however, is hampered by the pre-ethereality of the default reasoning³⁴, the impossibility of a comprehensive classification of all possible solutions to a problem³⁵, the large volume of common-sense facts³⁶, and the informal nature of some well-known knowledge³⁷.

In addition to setting goals and seeking optimal solutions, artificial intelligence systems can plan the results of their actions. Based on the initial data of the problem, the description of the desired goals and the set of possible actions to achieve the goal, the planning intelligent agents draw up an action scheme that is guaranteed to lead to an effective state containing the desired goals. The action scheme of today's artificial intelligence systems is based on techniques from:

- Classical planning - languages and algorithms for forward chaining state-space search, heuristic search, backward relevant-states search, and partial-order planning.
- Temporal planning - critical path method, minimum slack.
- Hierarchical planning - hierarchical task networks planning methods.
- Probabilistic planning - methods of dynamic programming, reinforcing training or combinatorial optimization of multiple iterative trial and error processes.
- Multi-agent planning - in environments with multiple agents, each designed with its own purpose, the ultimate common goal can be achieved through distributed, cooperative, competitive, or contested actions among the participating units.

Artificial intelligence systems are learning to solve current and potentially possible problems through a series of empirical observations. The learning process in many tasks cannot always be designed in advance (not all possible problems can be foreseen, not all changes in the environment can be predicted, it is not known how the behaviour of learning agents should be programmed, etc.) and can be implemented in a variety of forms, depending on ① what will be learned, ② what is already known, ③ how these data are presented, and ④ how the feedback in the agent is realized (Figure 1.8):

³⁴ Much of human knowledge takes the form of well-known postulates. For example, when it comes to a bird, people usually imagine an animal the size of a fist chirping and flying. However, these characteristics do not describe all existing bird species.

³⁵ According to John McCarthy et al. (McCarthy & Hayes, 1969) for each well-known rule there are several exceptions - the rules of abstract logic stipulate that almost nothing is just truth or falsehood, which greatly increases the number of potential solutions.

³⁶ Creating a base with all the common-sense knowledge in the world (such as the Cyc solution, for example) requires manual sequential design of all individual ontological concepts.

³⁷ Much of human knowledge cannot be verbally expressed through formal facts or rules. The representation of informal knowledge resulting from unconscious and intuitive human reasoning is still a difficult task for artificial intelligence systems.

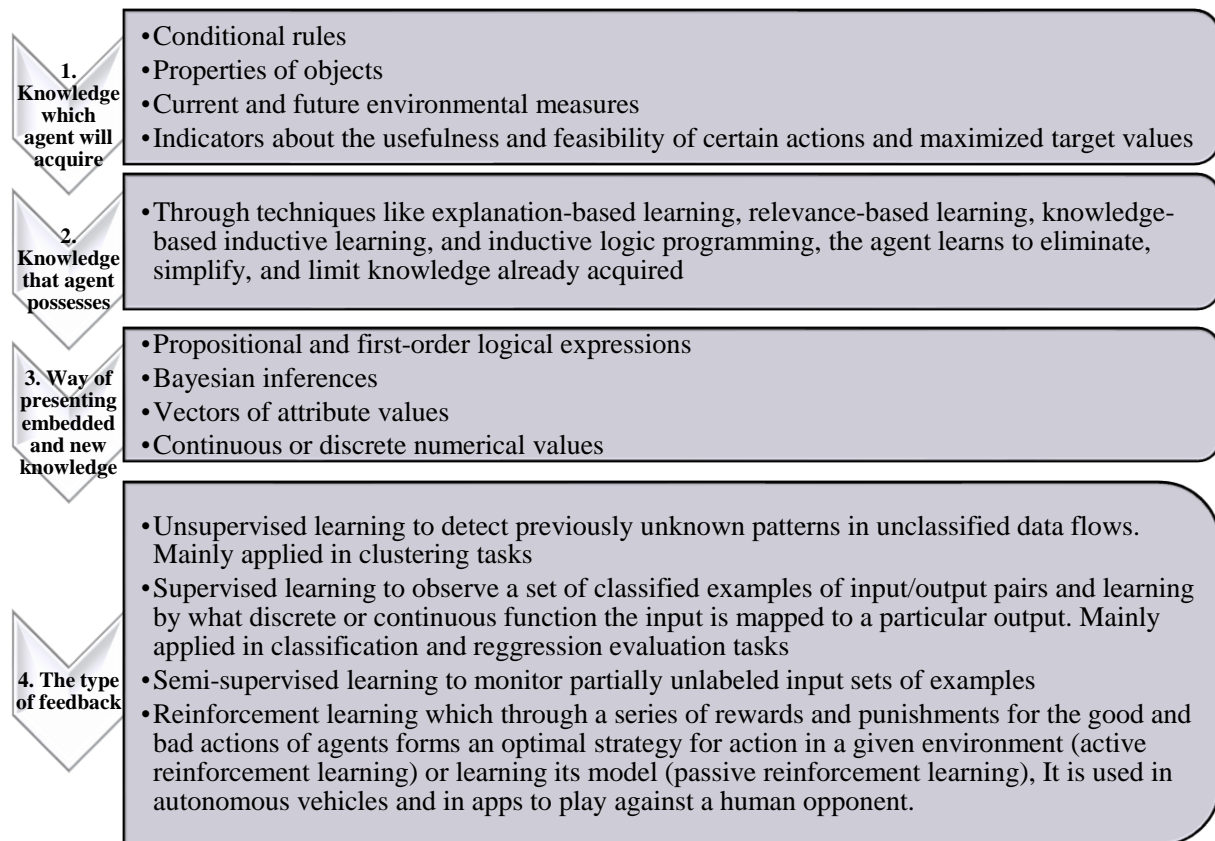


Figure 1.8. Features of the artificial intelligence systems learning process.

Source: author's illustration.

The successful development of artificial general intelligence solutions in the economic, social, medical, scientific, financial, and military spheres requires algorithmic training on huge sets of data.

The unique human ability to speak and express in writing through a particular language is implemented in artificial intelligence systems as opportunities for:

① Understanding the language patterns of natural languages through empirical study of human behaviour and learning certain grammatical and semantic rules. To predict the probability distribution of expressions in a language, intelligent agents make a lexical and syntactic analysis of the structure of phrases in it and interpret (semantically compose the meanings of individual sub-phrases and specify the most likely contextual, mental, linguistic, and acoustic meaning of a phrase) the recognized language patterns using lexicons parse trees, chart parsers, dictionaries, and corpora. Natural language understanding capabilities are built into artificial intelligence systems for machine translation and speech recognition.

② Processing of written sources (texts in newspapers and magazines, web pages, etc.) in the direction of classifying the text in them (this is how the artificial intelligence systems works for language identification, genre classification, sentiment analysis and detection of spam e-mails), retrieval of information (which is the main purpose of information search engines on the World Wide Web) and information extraction (a typical task of this kind is the extraction of results, corresponding to different syntactic and semantic templates). Modern

statistical approaches to natural language understanding and processing represent relatively accurately the syntactic structure of entire textual paragraphs or pages, but not the semantics needed to classify isolated sentences and describe common sense knowledge.

The emulation of biological abilities to perceive the real world and to explain the actions taken requires execution of a significant number of complex calculations by artificial intelligence systems. Information from their surroundings is obtained in the form of different types of signals (visual, acoustic, tactile, olfactory, radio, infrared, locational, wireless, reflective and radar), which are registered through the sensors and interpreted by the intelligent agent's program through a specific sensor model for:

① Visual perception of visible objects from the environment (object model) and representation of physical, geometric, and statistical processes in it (rendering model). Data-generating visual observations, on the basis of which agents manipulate, move in, and recognize their surroundings, are implemented in the artificial intelligence systems using several approaches: image formation, recognition of differences in objects and reconstructing a geometric model of the environment.

The images formed by extraction of geometric and physical descriptive characteristics (light dispersion, blurring, scale, light intensity, reflection, colouring, illumination, shading, colour constancy, etc.) are processed by the artificial intelligence systems in the direction of detection of edges, analysis of textures, calculation of optical flow and image segmentation of regions of similar pixels and super-pixels.

To recognize differences in the appearance of perceived objects, artificial intelligence systems apply the following types of templates:

- Abstract templates with known characteristics for the colouring, illumination, and orientation of object components, suitable for objects with few variations in the view (widely applied in facial recognition solutions – Figure 1.9).

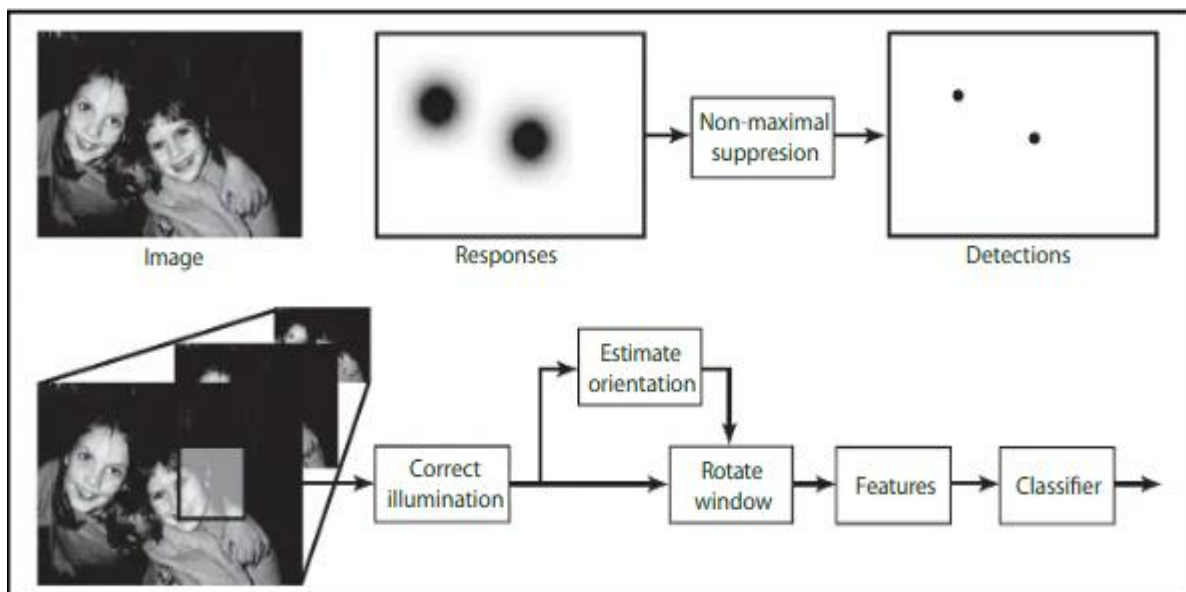


Figure 1.9. Architecture of a facial recognition system classifying images by key raster characteristics.

Source: Russel, S., & Norvig, P. (2016). *Artificial Intelligence: A modern approach (3rd ed.)*. Essex: Pearson Education Inc. p. 944.

- Complex templates that consider the foreshortening, aspect, occlusion, rotation and deformations at different positioning and scaling of objects, which can be described with different characteristics for size, colour, shape, etc.
- Spatial orientation templates using methods such as gradient histogram, scale transformation function, etc. Self-driving cars detect pedestrians through such templates.

Artificial intelligence systems reconstruct a geometric model of the environment from one or more images by triangulating different positions of the subject's capture or using physical characteristics from the background behind it. The three-dimensional information in the images is obtained by analysing various visual signals measuring the shooting angle, proximity, depth, distance, shading, contour, geometric shape, and structure of the objects in them.

② Auditory perception and processing of audio signals. In many smartphones, voice translators and cars, some form of machine hearing is used, which allows selective focus on specific sound from the surrounding environment. The ability to segment different streams of simultaneously incoming acoustic signals is implemented in artificial intelligence systems for sound search, genre recognition, acoustic monitoring, music transcription, music improvisation, emotion recognition in audio data, etc.

③ Haptic perception of mechanical and sensory stimuli from the surrounding environment (pressure, weight, friction, pain, temperature changes, surface changes, etc.). An example of successful emulation of the sense of touch and elasticity is the built-in tactile sensors in smartphone screens, robots, computer hardware and security systems.

④ Olfactory perception and classification of air-propagated components in the surrounding environment. 'Electronic nose' artificial intelligence systems are applied in the tasks of quality control of manufactured food, medical diagnosis, detection of narcotic, explosive and other illegal substances, response to climate disasters and environmental monitoring.

The ability to move and manipulate objects from the environment is realized through intelligent physical agents that can replace people in performing various activities, reproducing their actions. In the process of performing the tasks, robotic artificial intelligence systems operate with varying degrees of autonomy from human intervention. The multifunctionality of the application of robots in different manufacturing and health hazard activities is conditioned by the complexity of their physical construction (Figure 1.10):

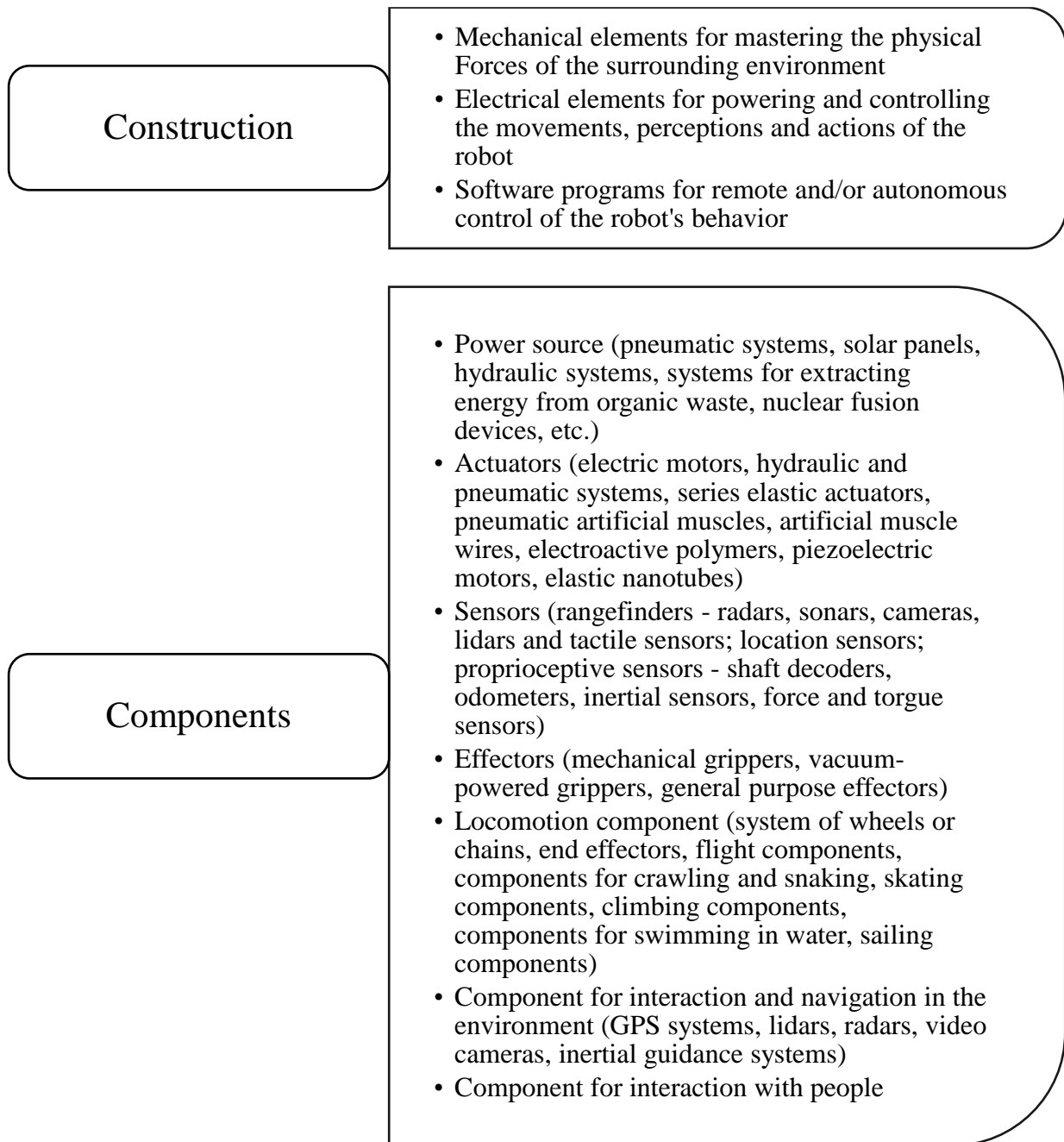


Figure 1.10. Physical components in robot's construction.

Source: author's illustration.

Robots constructed in the form of manipulators (robotic arms), mobile robots or mobile manipulators (humanoid robots) possess the possibility to:

① Perceives and evaluates likely environmental states (i.e., matching a particular sensory model to a motion pattern) through localization techniques, mapping, measuring various signals in it and machine learning.

② Plan their movements in their configuration space, each point at which it sets the location and orientation of the robot and the angle of its joints.

③ Perform controlled kinetic, differential, reactive or deliberate movement on a planned search algorithm trajectory.

④ Learn in a supervised and reinforced way that creates new skills for autonomous environmental research, social interaction with humans and improving robot behaviour.

Modern robots are developed according to the principles of subsumption, three-layer, or pipeline software architectures. The possibility of running many processes in parallel makes the latter suitable for self-driving cars (Figure 1.11), which combine methods of reactive control and of deliberate traffic planning.

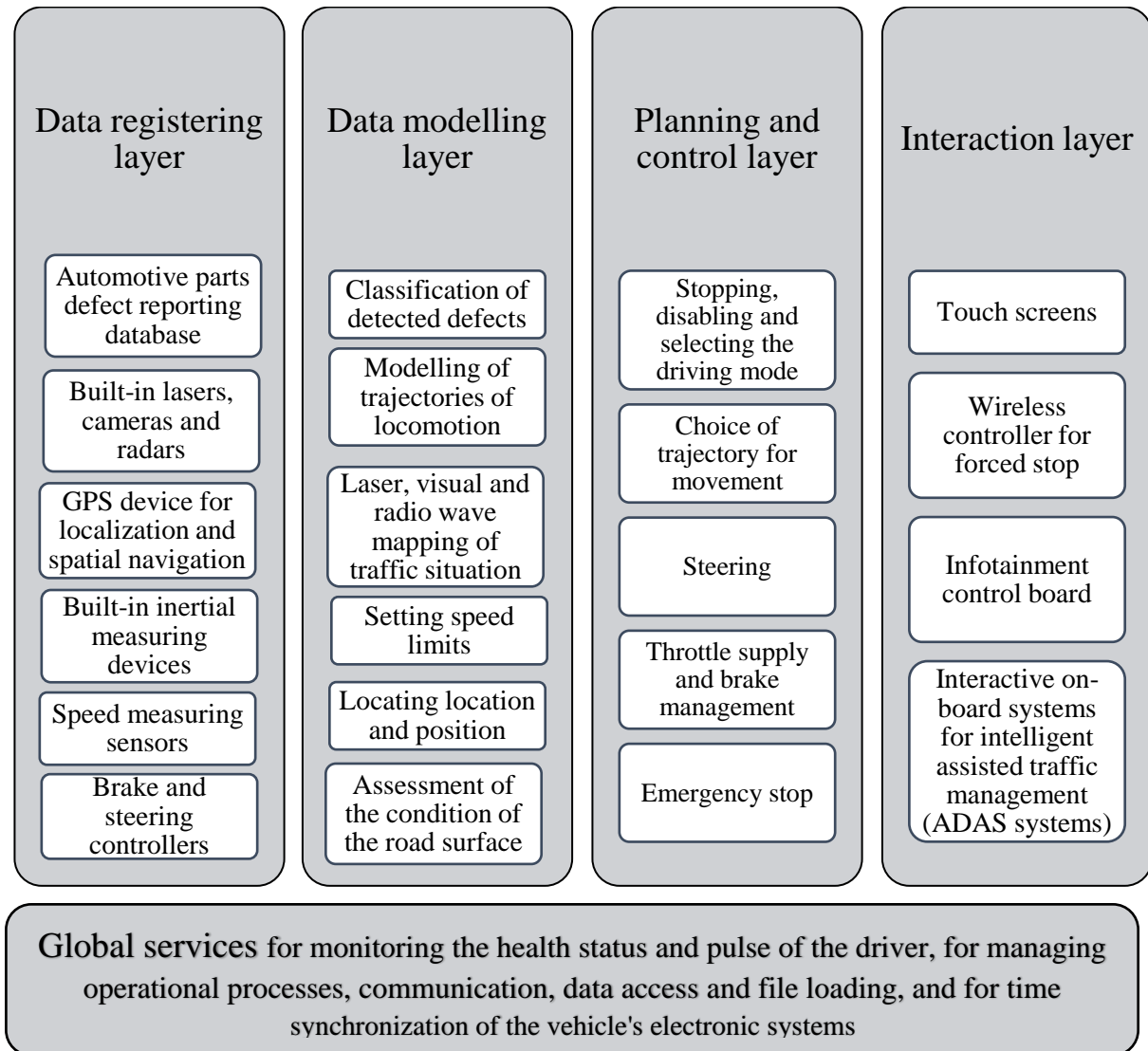


Figure 1.11. Pipeline software architecture of a self-driving car.

Source: author's illustration.

To be able to mediate the interaction with other intelligent agents, including humans, artificial intelligence systems emulate emotional and social skills to distinguish and process mental states, imitate feelings³⁸ and expressions and understand the motives for decisions. Social agents are machines with a certain level of emotional intelligence and empathy that can interpret emotions and adapt their own behaviours in response (Picard, 1995).

³⁸ Some virtual assistants, for example, are deliberately programmed to talk colloquially and joke, which gives users an unrealistic view of their real intelligence.

Artificial intelligence systems record data on the emotional state or behaviour of environmental agents through passive sensor devices (video cameras, microphones, thermometers, sphygmomanometers, or electroplated electrodes). The recognition of a specific emotional model from the collected data is realized using machine learning techniques and is applied in the tasks of speech recognition, natural language processing, facial expression recognition, gesture recognition, multimodal sentiment analysis (classifying affects in videotaped objects).

The human ability to create is implemented in artificial intelligence systems as an opportunity to create or display various works of art - images, sounds, animations, videos, literary works, musical works, video games, websites, visual algorithms, dance performances or exhibitions. Many traditional arts integrate digital technologies (for example, the artist can combine watercolour painting with visual algorithms and other digital techniques), leading to blurring the boundaries between classical and computer-generated creativity³⁹.

The listed intelligent capabilities of artificial narrow intelligence systems, implemented in part or in full in existing and future solutions of this kind, can be assessed by common benchmarks from Game theory (many games have a large base of professional players and established competitive rating systems) and specific tests to measure the degree of emulated intelligence.

The successful solution of various types of artificial intelligence systems tasks requires the implementation of opportunities for:

- Optimal performance of the task – artificial intelligence systems that can defeat a person in games of sea chess, related fours, checkers, Rubik's Cube, or poker.
- Super-human performance of the task– artificial intelligence systems that can play scrabble, backgammon, chess, go and solve quizzes.
- High-human performance of the task – artificial intelligence systems that solve crossword puzzles, play bridge and online video games.
- Par-human performance of the task – artificial intelligence systems for optical recognition of special characters, for image classification and for handwriting recognition.
- Sub-human (computational) performance of the task – lower than the human level of precision are demonstrated by artificial intelligence systems for object, face, and speech recognition, for visual answering of questions, for robotic movement, for explanation of medical diagnoses, for natural language processing.

A comparison between the degree of intelligence of artificial and natural systems can be made based on the following tests:

① A total Turing test requires a full-fledged artificial intelligence system to have capabilities for natural language processing, knowledge representation, automated reasoning, machine learning, computer vision, and robotic components.

³⁹ Developed in 2021 by the American research laboratory OpenAI and gained great popularity model for deep learning DALL-E 2 can generate images (including anthropomorphized versions of animals and objects) from text descriptions (prompts), combine unrelated concepts in plausible ways, render text and transform existing images.

② Subject matter expert Turing tests (Feigenbaum tests) - variations on the original Turing test assessing the work of experts in Chemistry and Marketing in solving problems of speech recognition, natural language processing and visual perception.

③ Completely automated public Turing test to tell computers and humans apart (CAPTCHA), verifying users of web services by solving tasks impossible to decrypt from a computer (most often reproduction of distorted letters, symbols and numbers and recognition of objects in images).

④ A test of universal intelligence aimed at comparing the current and potential possibilities of solving common problems of machines, humans, animals and even aliens (Hernandez-Orallo & Dowe, 2010).

1.4. Application of the Artificial Narrow Intelligence Systems

Detailed data on the global state of publication, research, studying, practical application, ethical issues, economic benefits, AI policies and regulations are cited and visualized in the Stanford Institute for Human-Centered Artificial Intelligence's fifth annual report. (Zhang, et al., 2022). The second chapter of the document quantifies (through common benchmarks tests and prize challenges) the technical capabilities of the most effective artificial intelligence systems in narrow domains such as computer perception of images and videos, natural language and speech understanding, recommendation making, reinforcement learning, computer hardware and robotics.

On the basis on the data in the report cited above and other theoretical studies, in the following paragraphs we summarize the application of modern artificial narrow intelligence systems in several major primary, secondary, tertiary, and quaternary economic sectors (Kenessey, 2012).

The agriculture sector, faced with humanity's ever-increasing need for more efficient yield and crop technologies, is deploying artificial intelligence systems with the ability to intelligently refine the following tasks:

- forecasting the ripening and picking time of a specific crop.
- Monitoring crops and soil condition.
- Use of statistical techniques for data mining, predictive modelling, machine learning, simulation, and optimization.
- Automation of greenhouse production.
- Robotic carrying out of agricultural activities (GPS localization of areas for fertilization and tillage, weeding by herbicides or lasers, harvesting, etc.).
- Recording the status of field biomass through images from drones and satellites, creating contour maps (Appendix 1), tracking water flows, and determining the sowing rate (Five technologies changing agriculture, 2016).

Despite ever-growing concerns that smart machines could take away people's jobs and cause social and moral problems, the industrial sector has a positive attitude towards artificial intelligence technologies, considers the transition to digital transformation an irreversible process and takes into account the economic benefits of industrial automation activities. The integration of intelligent design, manufacturing, distribution, and product lifecycle

management capabilities in industrial artificial intelligence systems (Appendix 2 and Appendix 3) mediated their application in the activities of:

- Increasing the user value of products in terms of efficiency, reliability, safety, and longevity.
- Improving the productivity of the production process through robotization, ensuring continuity in performing repetitive, dangerous or "humiliating" human activities, reducing operating costs in the long term, reducing the number of accidents at work, implementing distributed production systems.
- Knowledge discovery⁴⁰;
- Predictive analysis and preventive maintenance of the condition of industrial equipment by machine learning techniques.

The tertiary economic transportation sector is characterised by the availability of complex transport systems consisting of a large number of participants and vehicles for the air, ground, underground and water movement of persons and cargo. Artificial intelligence systems for transportation are expected to be safe, efficient, reliable, autonomous and carbon friendly to the environment and human settlements.

In the field of air transportation, the following artificial intelligence technologies are applied:

- Expert systems – to conduct flight simulation training, to support the implementation of piloting manoeuvres, to perform symbolic processing of simulation data, to train air traffic staff, to diagnose aircraft turbines, etc.
- Speech recognition software and artificial neural networks, with the help of which air traffic controllers train computer simulated piloting programs through voice instructions.
- Aircraft Conceptual Design Programs (for example AIDA - Artificial Intelligence supported Design of Aircraft).
- Software for flight control of damaged aircraft, compensating for the operation of defective components and allowing reaching a safe landing zone (Tomayko, 2003).
- Aircraft structural integrity management systems that process and interpret aircraft sensor data.
- Intelligent Autopilot Systems⁴¹ for carrying out machine learning of the autopilot system in the aircraft to deal with emergencies (bad weather, turbulence, or mechanical failure).

Thanks to the use of virtual development and testing tools, many manufacturers in the field of Information Technology and Automotive have created training algorithms for real-world driving automation, controllers based on the principles of fuzzy logic and ADAS systems for self-parking, advanced cruise control, reduction of stay, energy consumption and

⁴⁰ In critical tasks such as air traffic safety, for example, proactive flight risk management is implemented through parallel analysis of parametric data and flight text reports to detect anomalies and their relationship to specific causal factors. Understanding the origin of certain flight errors helps predict and prevent future incidents of a similar nature.

⁴¹ Intelligent Autopilot Systems combine the principles of apprenticeship learning (a form of supervised learning in which the system learns by observing the actions of human experts in the performance of certain tasks) and behavioural cloning to train the aircraft's autopilot system to perform certain manoeuvres.

harmful vehicle emissions, etc. Modern applied developments in terrestrial vehicles are aimed at achieving the third⁴² (Appendix 4), fourth and fifth levels of self-driving.

Currently, more than thirty IT and automobile companies are racing to create automated technologies for controlled vehicle braking, lane changing, collision prevention, navigation, and mapping. The creation of ground transport vehicles with a maximum level of self-driving requires the integration of several hardware-software systems for:

- Mechanical, voice and tactile perception of vehicle control signals.
- Operational control of the processes of navigation, localization, mapping, routing, sensory perception and tracking of the movement of the vehicle.
- Homogenisation of information from different sources in the environment to facilitate its transmission, storage, and processing by the vehicle's operating systems.
- Advanced communication system for sharing information with nearby self-driving vehicles⁴³. Connecting with other road participants and objects makes it possible to collect data that can be used to develop new autonomy features or improve mobility and safe driving capabilities.
- Software reprogramming at the initiative of the supplier or the vehicle machine learning system.
- Visualization of informational and entertainment data and update of the sound, visual and video content perceived by the environment. The effective movement of self-driving vehicles requires both pre-programmed geographical maps with the routes and peculiarities of the surrounding environment in the respective area, as well as the installation of devices to adequately respond to constantly changing road conditions.
- Safety. Some self-driving vehicles are not equipped with steering wheels or brake pedals, which requires the implementation of other algorithms to maintain passenger safety and deal with high-risk situations on the road.

In the Commerce sector (including online), artificial intelligence systems are applied for:

① Forecasting, pattern reconditioning, and consumer behaviour analysing. The use of artificial neural networks allows creation of personalized ads⁴⁴ (Appendix 5), guidance the process of selecting products and services⁴⁵, influencing over shopping decisions,

⁴² The world's first system with a third level of SAE autonomy classification certified for use in a production vehicle, Mercedes-Benz's Drive Pilot has already been approved for use on the roads of Germany and Nevada in the United States.

⁴³ Vehicular communication is an essential component of intelligent transport systems that are built as computer networks, communication nodes in which are vehicles and roadside units. The cooperative approach of interaction improves the efficiency of movement of road participants and reduces the number of traffic jams and accidents on the road. For now, similar experiments are mainly done in cargo transport, where the behavior of semi-autonomous and platoons of self-driving trucks is tested.

⁴⁴ The prediction and generalization of the information of users in order to offer personal promotions or automatically create customer profiles is made on the basis of their digital fingerprints on the web.

⁴⁵ The recommendation systems of Netflix, Amazon, YouTube, Walmart, Twitter, etc. generate audio and video playlists, product preferences from online stores or social media platforms and links to free web-sharing services.

understanding the motives for a particular purchase, and building lasting loyal relationships with the brand.

② Automation and increasing the efficiency of marketing processes. Traditional marketing tasks such as customer segmentation, marketing campaign management, product promotion and distribution are now performed automatically (in Amazon and Ocado warehouses, for example, multiple routine, repetitive and mechanical activities are robotic) using comprehensive intelligent monitoring and control systems.

③ Making management marketing decisions. The complex processes of developing new products and choosing the right mix of marketing tools today are supported by expert systems (MARKEX, BRANDFRAME, etc.), with the help of which brand managers can analyse market trends, make plans, reduce information overload, make joint decisions, identify key brand attributes, select retail channels, compare competing brands, prepare the budget for advertising, etc.

④ Analysis of social networks. The collection of social knowledge through techniques for emulating socio-based agents and for extracting data⁴⁶ can support the understanding of a market and of its segments. Modern search engines incorporate intelligent algorithms for collecting information about user-visited hyperlinks and search tools for web communities (PageRank, HITS). Such analysis helps to identify important and trendsetting participants or locations on the WWW and to take a socially responsible marketing approach in the areas of content creation, consumer intelligence, customer service, influencer marketing, content optimization and competitive intelligence (Ellett, 2017).

In the Tourism sector (mostly in the field of accommodation services) artificial intelligence systems are used to reduce the workload of staff, increase the efficiency of the activities performed, reduce duplicate tasks, analyse trends in travel, interact with guests and anticipate their needs. Hotels deploy chatbots, virtual voice assistants, serving robots (Appendix 6) and mobile applications with predictive analytics and real-time geo-location capabilities (Bhattacharjee, Seeley, & Seitzman, 2017).

The Finance sector approbates the artificial intelligence systems for:

- Prediction of declines in stock indices.
- Creation of financial plans.
- Detection of credit and debit card fraud and suspicious monetary obligations or claims. The use of anti-money laundering applications, for example (see Appendix 7), reduces the cost and time of investigating such crimes, minimizes false reports of such activities, and improves the detection of complex washing patterns, schemes, and connections (Boychev, 2021, p. 100).
- Assessment of the credit risk of insolvency of consumers and persons with limited credit data.
- Estimation of individual supply and demand curves and customised pricing for each user.

⁴⁶ The search and analysis of published on social networks information about the opinions, emotions and attitudes of people is realized through intelligent automated systems that help draw conclusions about the consumer attitude towards a product or service, filter the false reviews published on the Web and create psychologically targeted to certain socio-demographic groups advertising campaigns.

- Organization of financial operations in the non-working part of the day in the current time zone.
- Property management.
- Performing accounting and auditing operations on large data sets.
- Reducing financial crime by monitoring user behaviour patterns for strange changes or anomalies.
- Investment management (supporting investment decision-making, offering customized financial products to manage personal wealth, assessing the correlation between global events and their impact on asset prices, and more).
- personal finance management through specialized applications to support the process of optimizing savings and costs and through robo-consultants⁴⁷.
- Online algorithmic trading⁴⁸ and portfolio management of financial instruments.

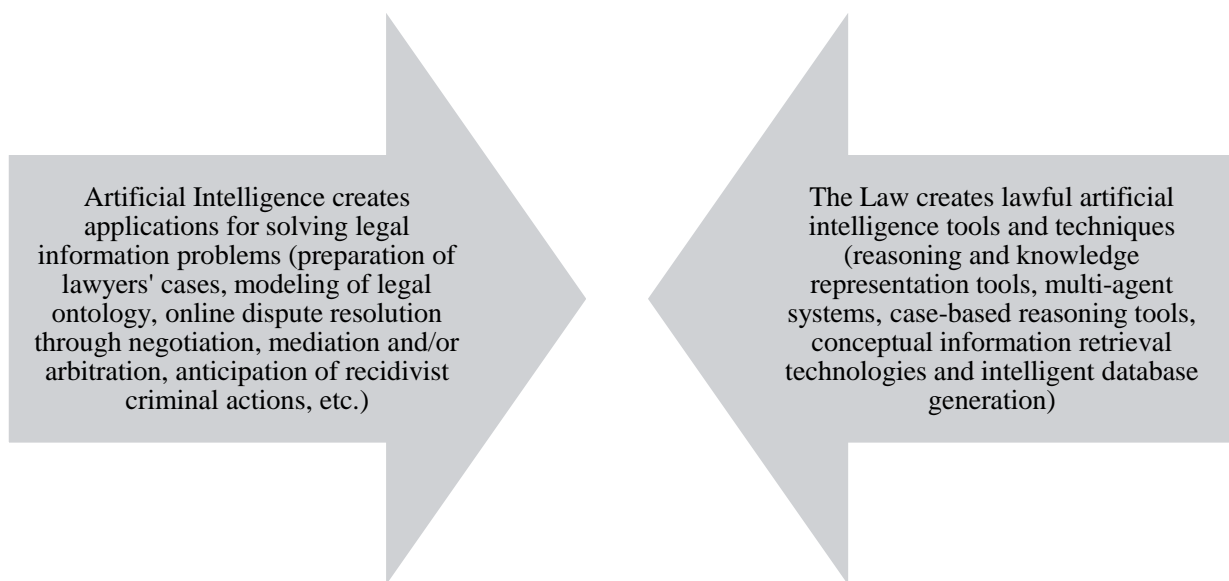


Figure 1.12. Interrelation between the fields of Artificial Intelligence and Law.

Source: author's illustration.

The interaction between artificial intelligence and the Law sector is two-sided (Figure 1.12). Artificial intelligence systems are used in various legal professions in solving a wide range of issues concerning:

- Formal models of legal reasoning.
- Computational models of argumentation and decision-making.
- Computational models of evidential reasoning.
- Legal reasoning in multiagent systems.
- Executable models of legislation.
- Automatic classification and summarization of legal texts.

⁴⁷ Robo-consultants provide financial advice on portfolio management and create an individual financial portfolio tailored to investment objectives and the level of risk tolerance of clients.

⁴⁸ In a 2001 algorithmic commodity trading competition, for example, six IBM softbots beat a team of six traders, earning 7% more than people.

- Automatic extraction of information and text from legal databases.
- Machine learning and electronic data mining from legal proceedings, government investigations, or open access to information lawsuits.
- Conceptual or model-based discovery of legal information (Appendix 8).
- Use of lawbots to automate secondary and repetitive legal tasks.
- Risk assessment, pricing, and timing of legal proceedings.

The implementation of artificial intelligence systems in the Education sector can reduce anxiety and stress caused by learning in a laboratory environment or interacting with educators (Sears, 2018). Intelligent tutor systems (Algebra Tutor, SQL-Tutor, EER-Tutor, Mathematics Tutor, eTeacher, etc.) possess possibilities for personalised learner support through lessons, case studies and games and for receiving immediate feedback. The use of social robots⁴⁹ for training pre-schoolers or learning certain skills (speaking a foreign language, for example) has the potential to significantly reduce wage costs in educational institutions (Appendix 9).

In the healthcare sector, complex machine learning and deep learning algorithms are applied, capable of collecting and processing information, recognizing behavioural patterns, creating their own logical rules, analysing the causal relationship between prevention or treatment techniques and the indicators of treated individuals, and drawing approximate conclusions about the status of patients. Medical artificial intelligence systems are used both in medical institutions and hospitals (to save costs, improve patient satisfaction and meet the needs of staff and workers) and in business organizations (to improve the efficiency of business processes, reduce patient length of stay and optimize staff workload).

The application of artificial intelligence systems in healthcare is in many directions:

- Diseases diagnosing.
- Development of treatment protocols⁵⁰.
- Medicines creating.
- Personalised medical care by selecting the most effective drug combinations for each individual patient.
- Monitoring and care of patients through clinical medical decision support systems.
- Robotic surgery with autonomous operating systems.
- Computer-aided interpretation of medical images.
- Analysis of heart sound.
- Caring for older people with companion robots.
- Extracting useful information from medical records.
- Planning the design of treatments.
- Medical consulting (Appendix 10).
- Supporting routine and repetitive medical activities.

⁴⁹ Disadvantage of educational artificial intelligence systems is the probability of learners' loss of concentration and irrationality (which is not the behavior of intelligent agents) when solving problems. Although social robots give information of the same quality as their human counterparts, for now they are used primarily as learning tools rather than teachers.

⁵⁰ In 2016 in California, for example, with the help of artificial intelligence, a mathematical formula was developed for accurate dosing of immunosuppressants in people with transplanted organs.

- Using avatars instead of patients for clinical training.
- Predicting the likelihood of death in surgical procedures.
- Predicting the progression of AIDS disease.
- Safe performance of work tasks causing stress, overload, musculoskeletal injuries, and other diseases (e.g., implementation of chatbots to replace employees in call centres).

In the energy utilities sector, the efforts of AI researchers are primarily focused on automating the process of designing reliable power electronic converters⁵¹ to create devices with cost-effective subsequent maintenance, maximum durability, and resistance to damage (especially for geographic areas critical to the sector or country). Some telecom operators manage their employees' work schedule through the heuristic search method⁵².

In the field of social services, artificial intelligence systems are applied, supporting the processes of:

① Human resources management. The activity of recruitment specialists is already successfully automated by intelligent HR platforms (specialized software applying natural language processing technologies to extract and highlight key information from candidate documents), communication chatbots and robots for interviewing and assessing the skills and educational qualifications of applicants for a particular job position (Appendix 11).

② Job search (including the processes of online hiring, searching, applying for and changing jobs). Artificial intelligence applications collect information about skills, requirements and expected pay from candidates, offer the most suitable jobs, calculate what remuneration would be most appropriate for a particular job, and more.

④ Automated online assistance. Virtual chatbot assistants for online dialogue, represented graphically by avatars of a human or other figure and performing tasks or services based on commands and questions, are part of the websites of many business organizations⁵³ (Appendix 12). Voice-controlled software artificial intelligence systems receive test messages, interpret human speech, respond through synthesized voice, control other automated devices, reproduce multimedia information, manage electronic correspondence, to-do lists and calendars, search for information by capturing and/or uploading images, and more. These are integrated into various instant messaging applications and platforms and installed on personal computers, smartphones, smart speakers, home appliances, cars, and wearables.

In the Defence and National Security sector, machine learning algorithms and expert systems are applied in order to:

- Improving command and control in the performance of military missions involving vehicles with different levels of self-driving.
- Reconnaissance collection and analysis of logistical and cyber-information to detect and identify threats, mark the position of the enemy, intercept targets, etc.

⁵¹ These are electronic components with wide application in the field of renewable energy generation, energy storage, electric vehicle design and DC high-voltage power grids.

⁵² In the British BT Group, for example, such an application creates a schedule for work and rest for 20,000 employees.

⁵³ The use of softbots makes it possible to streamline communication with customers, reduce the cost of operating telephone systems and training staff and improve dialogue with ill-intentioned customers (Clark, 2016).

- Improving communication, coordination, integration, and interoperability in carrying out distributed firing attacks using networked combat vehicles, tanks and human or unmanned aerial squadrons (Slyusar, 2019).
- Conducting combat flight simulations.
- Choosing successful tactical solutions in air combat.
- Implementation of cybersecurity solutions protecting state information systems from various malicious attacks (Appendix 13).

The application of artificial intelligence technology in the Government sector increases the efficiency of administrative services provided, reduces the cost of paying front office employees and reduces the opportunities for corruption (Mehr, 2017). Artificial intelligence systems change the way several categories of public tasks are dealt with:

① Achieving public policy goals:

- Automation of the process of obtaining state regulated benefits.
- Provision of social services.
- Monitoring of social networks for public feedback on specific government policies.
- Adjudication of judgments on the waiver of a guarantee.
- Supporting the activities of revenue collection and immigration services.
- Classification of calls to the national emergency telephone system.
- Identification of risks for the purpose of deceiving the population
- Predicting crimes and recommending optimal police presence.
- Prediction of traffic congestion and road accidents.
- Prediction of maintenance needs of certain roads.
- Detection and prevention of the spread of diseases.
- Sorting of health insurance cases.
- Identification of violations of health regulations.
- Development of chatbots for health diagnostics.
- Delivering personalised school education.
- Marking of examination documents.
- Support for defence and national security, etc.

② Supporting the interaction with the public administration:

- Answering questions through virtual assistants (Appendix 14).
- Redirection of requests for administrative services to the relevant office.
- Automatically fill out forms.
- Assisting the search for documents.
- Scheduling meetings with administrative staff.

③ Other tasks - machine translation and preparation of documents.

Artificial intelligence systems are also actively applied in the Media (mainly electronic ones) and Culture sectors. We increasingly read about automated creation and distribution of

publicist content (texts⁵⁴ (Appendix 15), news⁵⁵, statistical reports, personalized recaps, visualizations, articles⁵⁶ (including scientific ones), videos, social media posts, comments, recommendations, summaries, and stories) offering opportunities for personalized experiences at a user-appropriate time.

In recent years, social media has replaced television as a source of news for people of generations X and Z, giving rise to the creation of artificial intelligence systems handling audio-visual content. Thanks to computer vision technology applied to object, face or scene recognition tasks, films, TV programs, advertising videos and user-generated content (Appendix 16) are analysed in the direction of:

- Facilitating the search for audio-visual information in various electronic media.
- Creating a set of descriptive keywords for a specific audio-visual object.
- Compliance with content posting policies.
- Converting speech to text for archival and other purposes.
- Detection of logos, products and faces of celebrities to place them in advertisements.
- Creating animated faces and expressions of real people through deep artificial neural networks. The generation and distribution of deepfakes, originally conceived for humorous purposes, is now synonymous with false documents, cloned human voices, and video-replaced personalities. Countering this threat is also through artificial intelligence programs and is funded by multimillion government programs⁵⁷.

Artificial intelligence systems are already being applied to the creation of musical content in order to:

- Creation of therapeutic music for pain and stress relief based on natural human biorhythms.
- Composition of songs in a certain style and combination of styles after analysing huge databases of musical styles and optimization techniques for sound extraction (Google's MusicLM apps (Appendix 17) and Sony's Flow Machines, for example).
- Interactive composition of musical works through computer accompaniment of live performance by musicians.
- Distribution and control over the use of musical compositions, etc.

The production of artificial intelligence systems for education and entertainment, which began as early as the 1990s, today results in the creation of: robotic toys with built-in software for speech recognition, interactive dialogue and learning; computer-simulated board games (chess, Go, checkers, poker, etc.); video characters (bots) with responsive, adaptive

⁵⁴ The extremely popular chatbot at the end of 2022 ChatGPT writes complete and meaningful paragraphs, pages and fragments of software code.

⁵⁵ On <https://hanteonews.com/en/article/bot%20news> website, for example, news written by softbot is published.

⁵⁶ In 2016, a Japanese AI program co-authored a short novella that nearly won a literary award. (Olewitz, 2016).

⁵⁷ Within the framework of the European Horizon 2020 program, for example, the InVID Project has been initiated, and in the US such a program is IOGAN ACT: Identifying Outputs of Generative Adversarial Networks Act.

and intelligent behaviour, competing against which aims to improve the experience of participants in different games⁵⁸.

Over the years, numerous generating art content artificial intelligence applications have been developed - Disco Diffusion, DALL-E (1 and 2), Stable Diffusion, Imagen, and others⁵⁹ (Appendix 18). In addition to creating art, artificial intelligence systems are also used to analyse digitized collections of artistic works and to predict emotional reactions to such works⁶⁰.

The role of machine learning algorithms in computer-created art is popularized by the American Computer Machines Association (ACM), conferences on artistic topics and various exhibitions - 'Thinking Machines: Art and Design in the Computer Age, 1959-1989', 'Unhuman: Art in the Age of AI', 'Understanding AI' and 'Uncanny Values: Artificial Intelligence & you'.

⁵⁸ The targeting of bot actions in many modern video games is based on intelligent techniques for determining the critical path, forming a decision tree, data mining and generating procedures. For many researchers in the strand, however, 'game AI' is not a real artificial intelligence, but an advertising trick describing computer programs with sorting and matching capabilities, creating only the illusion of intelligent behavior.

⁵⁹ Generative artificial intelligence systems (autonomous hardware-software systems that can independently determine the characteristics of the content they create) apply mathematical models and algorithms simulating brushstrokes and other pictorial effects.

⁶⁰ A dataset with machine learning models that contain emotional reactions to visual art is available at <https://www.artemisdataset.org/>.

Chapter Two. Methods, Principles, and Algorithms of Artificial Narrow Intelligence Systems

2.1. Methods for Searching, Optimisation, and Classification

Problem-solving intelligent agents use a common search algorithm to extract information in the form of discrete or continuous values from stored data structures. The performance of classical search algorithms is evaluated by several indicators: speed of finding a solution⁶¹, optimality of the solution found, quality of the ranked returned results and computational complexity. Artificial intelligence systems use several basic search methods – searching in a space of states of the application domain, blind searching, heuristic searching and searching in an abstract space of states.

State space search is a method of considering possible solutions to a problem in pursuit of the desired solution. The space of possible solutions is usually depicted as a tree graph structure, whose nodes are the states of the problem, and the arcs are operators leading from one state to another. The search procedure, consisting of applying a sequence of operators that change the state of the data structure from an initial state to a target one, is performed in a straight, reverse, or simultaneous manner in both directions.

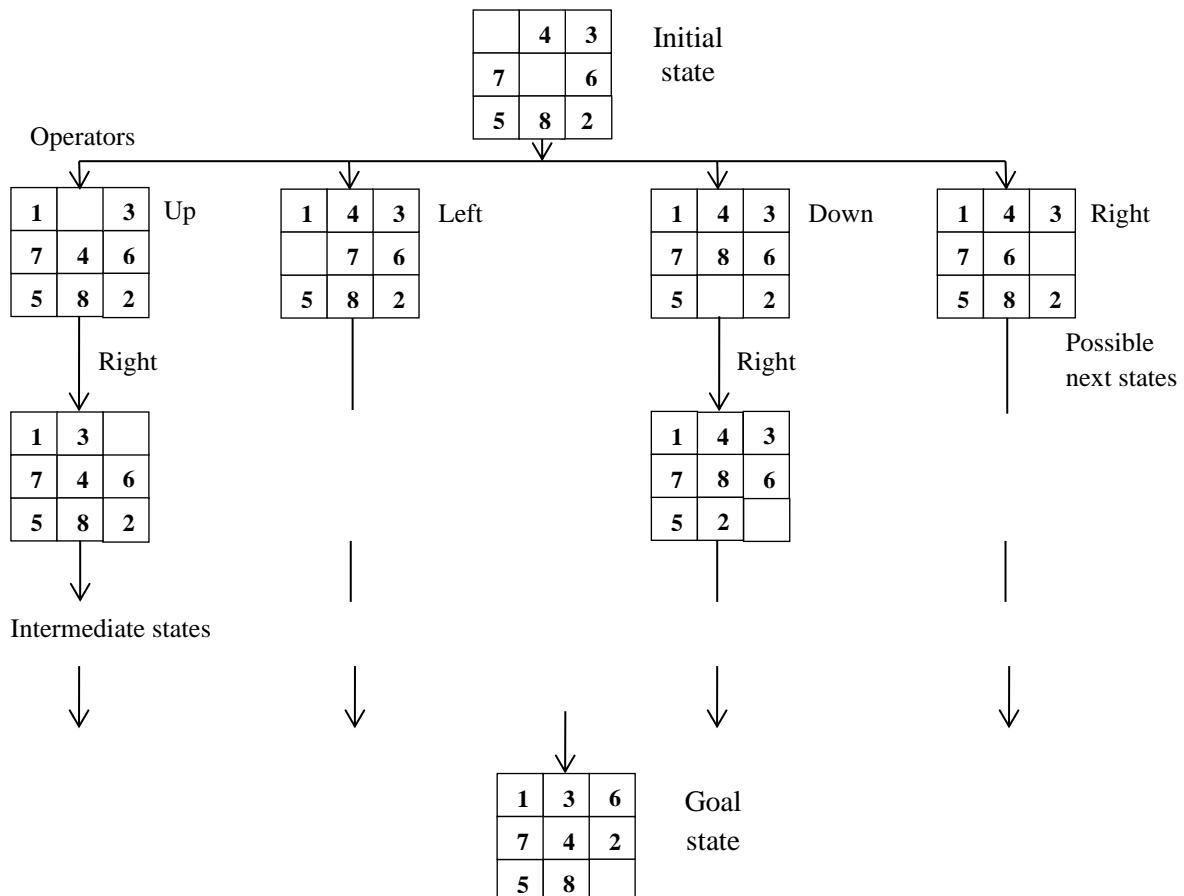


Figure 2.1. Solving the ‘8-puzzle’ toy problem through state space search.

Source: author's illustration.

⁶¹ Search trees, hash-maps and database indexes, for example, are specially designed structures for accelerated finding of solutions.

An example of using state space search method is the '8-puzzle' toy problem, in which a matrix with eight square cells, numbered from 1 to 8, and one empty cell must be displaced so that the cells are positioned in a predetermined order (Figure 2.1). The motion of the empty cell along the vertical or horizontal is described by the terms of the operators viewing the empty square as an object which can be moved in any of the four directions. Each of the four operators can change the matrix, giving rise to a new state. Certain operators may not be applicable to a given state (in the example if the empty cell is located at the end of the matrix and needs to be moved down or to the right). In the '8-puzzle' toy problem the search is performed in the opposite direction (from the desired final state to the initial condition) by applying reverse operators.

The solution search in graphically represented data structures is done in different ways. The blind search method systematically examines all the arcs of the graph in breadth or depth. The breadth-first search performs a gradual uniform descent along all possible states from first, second, third and so on levels (Figure 2.2 (a)). This way of simultaneously searching for several possible solutions requires artificial intelligence systems with a high storage capacity. The depth-first search "sinks" in the spatial graph - initially, nodes of one branch of the tree are successively considered, starting from a first-level state, then its derivative state, then the derivative of the derivative state, and so on, and then in the same way the nodes of the alternative branches of the tree are traversed (Figure 2.2 (b)).

Although both blind search methods can achieve the goal of the problem, solving complex problems with too many branches in the process of traversing the graph costs significant volumes of computational time and memory.

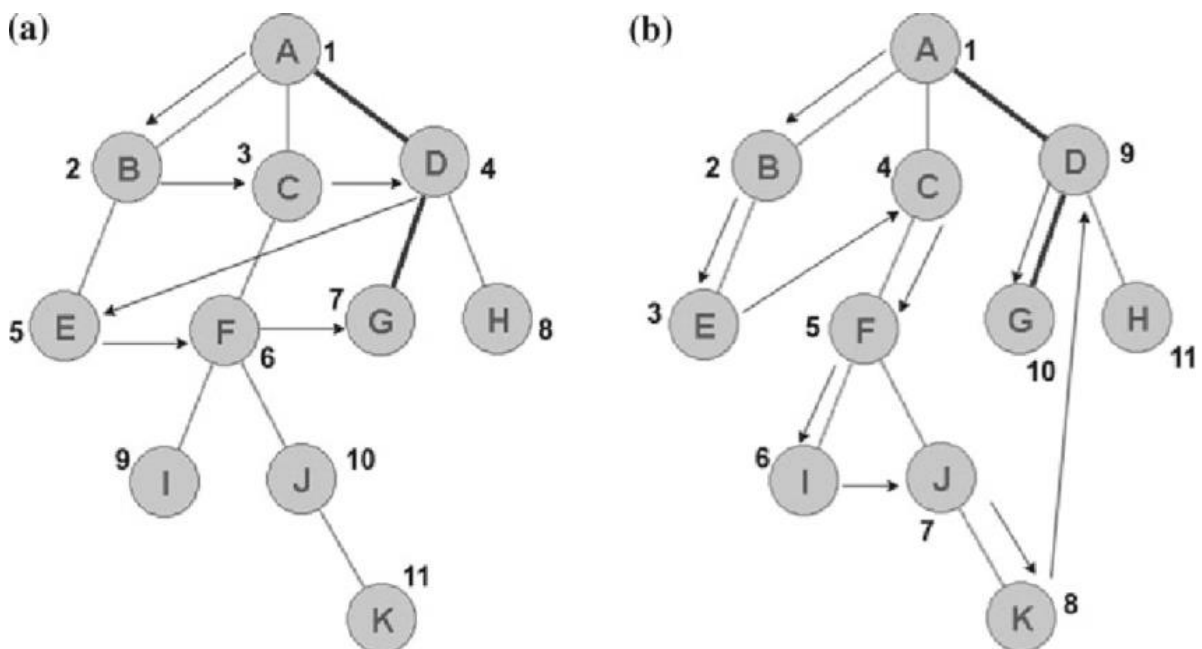


Figure 2.2. Blind search methods in breadth (a) and in depth (b).

Source: Khemali, C., Doshi, J., Duseja, J., Shan, K., Udmale, S., & Sambhe, V. (2019).
Solving Rubik's cube using graph theory: ICCI-2017. p. 308.

In order to make the problem-solving process more efficient and to reduce the volume of calculations, many artificial intelligence systems apply heuristic methods to select the most likely achievable goal in the smallest number of steps. The heuristic function builds a branched search algorithm and arranges the possible alternatives at each branching point of the graph. The solution process starts from the 'best' node and continues along the branch that is judged⁶² to be closest to the final goal. The evaluation function for measuring the computational costs along the selected branch depends on the application area of the problem being solved (the search for the shortest route between two points, for example, is most often measured by the magnitude of the distance between them (see Figure 2.3)).

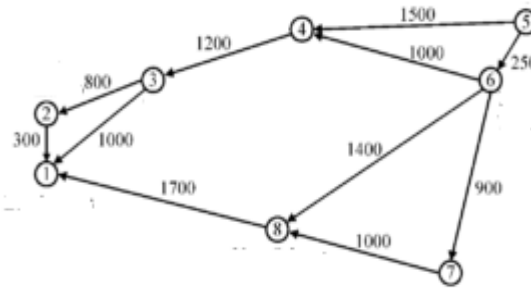


Figure 2.3. Heuristic search method of the fastest route between points 5 and 1.

Source: author's illustration.

Solving the problem of finding the shortest path does not require a detailed mapping of all possible paths between the two endpoints (unless a stop is planned at some intermediate point), but only of the main ones. To determine the optimal route, artificial intelligence systems sometimes use large-scale contour maps that are resultant abstractions of the detailed maps.

Searching in an abstract state space is a faster method because it works with consolidated steps from the source space. In the example (figure 2.3), if a specific path is chosen between points 5 and 1, the initial task is broken down into several smaller tasks (first to point 6, then to 8), which can be analysed by an intermediate level of abstraction. The hierarchical way of reasoning (the use of different levels of abstraction) does not require the description of all intermediate points in the search and speeds up the process of solving the problem by the artificial intelligence systems.

Sometimes the search in a state space is done using the procedure 'generation and verification'. It divides the process into two parts: generating the possible solutions; testing for prune decisions that do not satisfy certain constraints. The artificial intelligence system is complete if it is able to create every possible solution. It is without residual if each solution is generated only once. In many AI applications, the processes of generation and verification overlap over time.

It is important to divide knowledge between these two processes. Often the search is most effective if the maximum amount of knowledge is included in the generation system. A

⁶² The estimation of the space of states in an adequate way requires a considerable amount of symbolic computation. The heuristic way of solving focuses on the domain-specific information by reducing the number of branches of the graph.

very efficient procedure in this respect is the hierarchical one, by which hopeless solutions which are only partially determined can be pruned. When the partial description is rejected, the whole class of solutions that meet this description is excluded from the generation process. Through the hierarchical procedure, powerful rules are applied to reject inappropriate options at the initial stage of the search process.

The influence of Mathematics on the field of Artificial Intelligence is expressed in the approbation of several optimization methods – blind hill climbing, simulated annealing, beam search, evolutionary computations, random optimization, and swarm intelligence.

Blind hill climbing is an iterative mathematical algorithm in which a search begins with an arbitrary solution to a problem but is refined incrementally to the point where no further improvements can be made to the result.

Simulated annealing represents a probabilistic technique to approximate the optimal action of a function suitable in search spaces with discrete values and to find an approximate global maximum/minimum.

The beam search method is a heuristic algorithm that explores the decision tree by extending the most promising node to a limited set of alternatives and storing the best solutions.

Evolutionary computations are a set of global optimization algorithms inspired by biological evolution and soft computing that solve problems of a metaheuristic or stochastic nature by trial and error. In this method an initial set of possible solutions is created which is updated iteratively (Figure 2.4) - each new 'generation' of solutions is produced by removing the less desirable solutions and introducing small random changes.

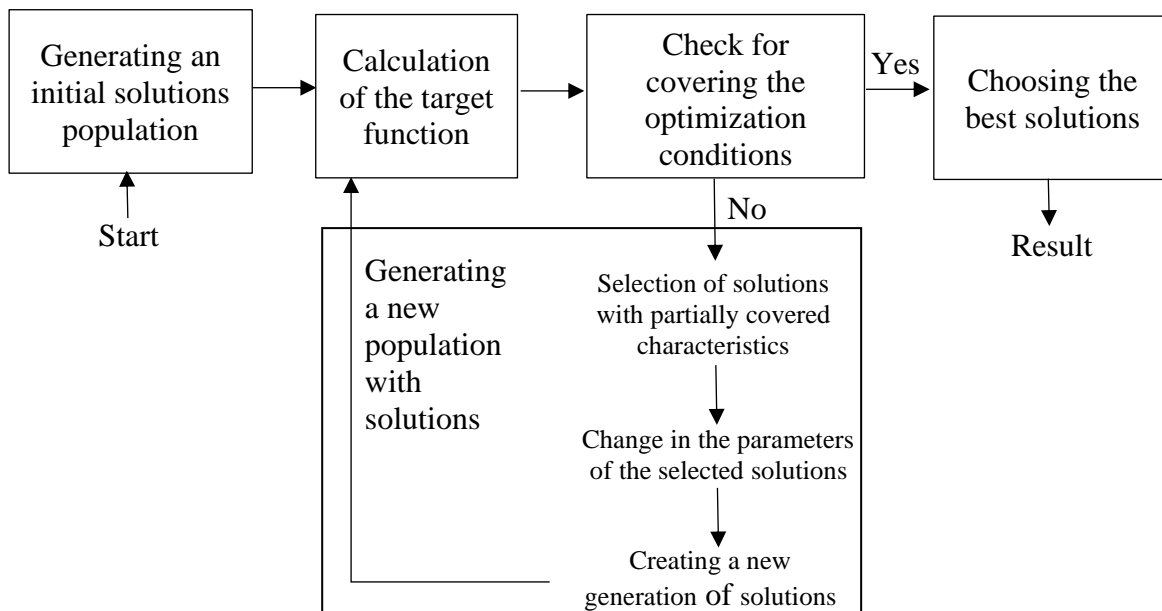


Figure 2.4. Algorithm of the evolutionary method of decision-making.

Source: author's illustration.

Random optimization (direct search, derivative-free search, or black box method) is a set of numerical optimization methods that can be used for tasks represented by continuous or differentiable functions.

Swarm intelligence algorithms are a computational method for searching distributed spaces that is based on the collective processes occurring in swarms or colonies of biological organisms. This method creates a 'population' of possible solutions, which are recalculated using simple formulas and moved in the search space according to their own best position and that of the entire population. When a better solution is found for any member of the population, the movement of the entire population changes in a coordinated manner.

In the field of supervised machine learning, different statistical methods, principles, and tools are applied, although the main purpose of the two areas differs significantly⁶³. In both strands, however, classification problems are solved to determine the category to which each new pattern should be assigned in a certain set of data. Classification algorithms use pattern matching and are selected according to the characteristics of the data in the set. There is no classifier with universal application for all problems, but the choice of a specific one can be made according to the indicators such as dataset size, patterns distribution by class, number of dimensions to analyse the data and degree of inaccuracy in the patterns. Classification methods, as we have already pointed out in paragraph 1.2., are also applied in the artificial intelligence systems for pattern recognition.

The classification algorithms for the different data sets can be trained by several statistical and machine learning methods – artificial neural networks, decision trees, k-nearest neighbour algorithm, support vector machines, Gaussian mixed model, and naïve Bayesian classifiers.

Artificial neural networks are computational systems inspired by biological neural networks that, without programming with specific rules, are trained to perform tasks on examples of data.

Decision trees represent decision support tools that use a hierarchical model to depict possible decisions and their possible consequences, including random events, resource costs, and utility. In essence, tree models are a structure of a discrete set of values, the leaves in which represent labelled classes of data, and the branches – specific characteristics of the classes. The assignment of continuous values (usually in the form of real numbers) by the target variable turns the tree into a regression one (see Figure 2.5.).

The k-nearest neighbour algorithm is a nonparametric method in which the input data contains the closest from the feature space learning examples. In this 'lazy' learning algorithm, the mathematical function is approximated only locally, and all calculations are postponed until the final calculation of the function.

⁶³ Statistics draw conclusions about the entire population by analyzing a sample of this data, while machine learning finds generalizable predictive patterns in the data.

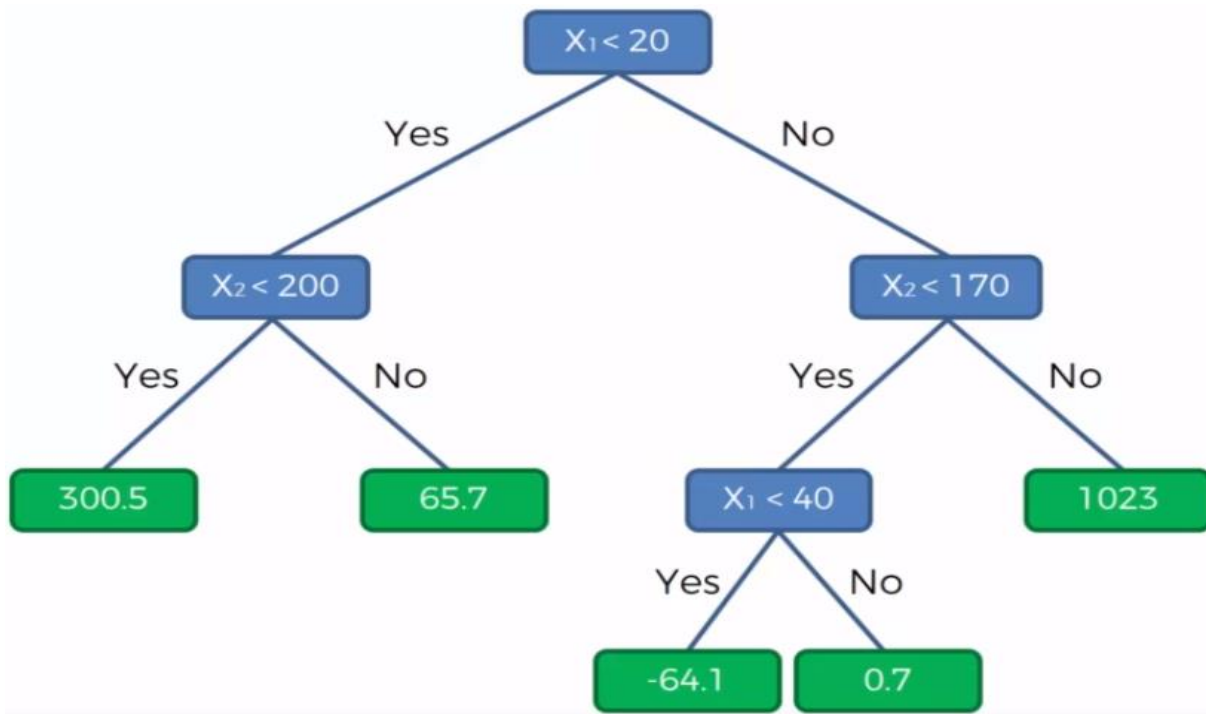


Figure 2.5. Decision tree to check the values of two variables.

Source: Girgin, S. (2019, May 22). Decision tree regression in 6 steps with Python. <https://medium.com/>.

Support vector machines (SVM) are a set of related supervised learning methods in which the learning algorithm builds a model that predicts the fall of new examples into one category or another. SVM training algorithms are mostly non probabilistic binary linear classifiers and can be applied in linear and nonlinear classification problems where implicit adjustment of input data to spaces with multidimensional characteristics is needed.

The Gaussian mixed model is a probabilistic model to represent the presence of sub-populations within the entire population, in which the vector of numerical values is extended to cover unknown parameters or multivariate normal distributions in the data.

Naïve Bayesian classifiers are a family of simple probabilistic classifiers that admit strong independence between the characteristics of the vector of numerical values.

2.2. Principles for Logical and Probabilistic Reasoning

Artificial narrow intelligence systems apply a number of logical principles in solving problems requiring opportunities for reasoning and presentation of knowledge, for planning, for training and for classification. The behaviour of rational thinking agents is programmed on the basis of concepts of several basic logical partitions – propositional logic, first-order logic, fuzzy logic, non-monotonic logic, default logic and circumscription.

Propositional logic is used to define simple (atomic) or complex statements called propositions, which can be True or False. In essence, it is a formal system for presenting statements about facts in a domain. The evaluation of the truthfulness of each statement is done according to certain inference rules and axioms, which create derived formulas (theorems) describing positive claims in the domain. The use of logical inference rules (substitution rule, Modus ponens, Modus tollens, concretization, deduction, resolution rule,

etc.) allows to derive new facts about the application domain by dynamically creating new symbolic structures and new well-defined formulas. Propositional logic does not deal with illogical objects, quantitative variables, and relations between facts.

First-order logic is an extension of propositional logic that expresses illogical facts about objects in a given domain, their quantitative properties, and their relationships with each other. In this most commonly used formal method of representing symbolic structures in computer programs, the facts of the application domain are expressed by:

- Atomic sentences - formulae consisting of a single predicate symbol and a list of terms.
- Complex sentences – well-defined formulas combining several atomic sentences connected by logical functions.
- Quantifiers - symbols to denote quantifying functions by means of which judgments characterizing the domain of truth of a given predicate can be constructed.

Fuzzy logic assigns a 'degree of truthfulness' between 0 and 1 for vague expressions that are difficult to refine as true or false from a linguistic point of view⁶⁴. The fuzzy models and sets (Zadeh, 1965; Klaua, 1965) are mathematical tools capable of recognizing, representing, manipulating, interpreting, and using obscure data and information. They are successfully applied in control systems because they allow to set vague rules for describing the facts of the domain, which can subsequently be numerically specified by the intelligent agent.

Unlike formal Boolean logic, where only simple one-order logical operations of the type AND, OR and NOT are performed, the statements in fuzzy logic can take on a large number of values⁶⁵, since variables with nonnumeric values are used to define facts (the element 'sex', for example, traditionally takes the values 'man' or 'woman'). Input numeric values are matched to certain fuzzy functions by Zadeh operators and hedges.

Non-monotonic logic is a formal logical system for drawing and presenting inaccurate conclusions about facts in the domain, in which reasoners make temporary assumptions that may change when new information is received. Learning new knowledge allows you to quickly reduce the set of known solutions to the problem. Non-monotonic reasoning plays a major role in the construction and planning tasks, where the space of possible solutions often turns out to be very large and it is impossible to predict the consequences of choosing any of them. After formulating an assumption, the intelligent agent checks whether it satisfies existing constraints and, if not, produces another assumption. As a result of considering the consequences to which the first assumption leads, new information about the described fact is created.

The default logic is a non-monotonic logic formalizing assumptions about facts from the domain with a previously known meaning (Reiter, 1980). Such assumptions are based on

⁶⁴ For example, a person may be short or tall or shorter or above average height or very tall; Braking a car does not necessarily mean applying the brakes, and the increase in speed is not due to the accelerator being applied alone.

⁶⁵ Each element in the fuzzy set is of a dual nature: if only absolutely false and absolutely true facts are used in formal logic, here they are semi-true - with some degree of certainty obtained by mixing the wrong and the true. The relationship between these two extremes determines which of them is stronger and whether the meaning of the element tends more towards the right or the wrong. Because of this property, fuzzy logic facilitates the description of real objects that rarely accept unambiguous or only two values.

uncontradictory common knowledge, and their results are valid until proven otherwise. Expressions in default reasoning are formed by using predicate logical relationships and rules to correlate certain prerequisites and consequences.

The circumscription of John McCarthy (McCarthy, 1980) formalizes the common-sense knowledge of the facts of the various domains and solves the classification problem formulated by him. The researcher extends first-order formal logic with fixed and varying predicates to minimize false facts that are supposed to be a lie.

When solving problems in specific areas of knowledge, artificial intelligence systems also apply principles of description logic, situational calculus, event calculus, fluent calculus, causal calculus, belief calculus, modal logic, and paraconsistent logic.

Since logical principles are difficult to apply to solving problems with numerous and contradictory options for action, and many tasks of reasoning, planning, learning, perceiving, and navigating require the intelligent agent to operate with incomplete, uncertain or inaccurate information, the field of Artificial Intelligence also appropriates mathematical principles for probabilistic reasoning - Bayesian networks and techniques in the field of Economics.

Bayesian networks are a statistical approach to calculating the degree of definiteness of claims about facts in the domain and to predict the probability that the causes of a consequence are an influencing factor. They are illustrated by a probabilistic graphical model representing a set of variables (observable quantities, latent variables, unknown parameters, or hypotheses⁶⁶) and their conditional dependencies, and are applied in the following intelligent tasks:

- Reasoning using Bayesian inference – an algorithm for drawing conclusions in which the probability of a hypothesis is updated in the presence of new information through the Bayes' theorem.
- Learning through the expectation-maximization algorithm – an iterative statistical method for finding the maximum likelihood or maximum value of parameters in statistical models dependent on unobserved latent variables.
- Planning the solution of problems through influence diagrams (Figure 2.6) – a generalization of Bayesian networks, in which a situation is presented in a compact graphical and mathematical form.
- Perception using dynamic Bayesian networks that link the various variables in the network model through approximate time steps. Dynamic Bayesian networks are used to filter, predict, smooth, and find explanations in streams of variable data (e.g., speech signals, digitally stored data, or protein sequences) of time-consuming processes. They were developed for the purpose of generally representing and deriving probabilistic inferences about arbitrary nonlinear and time-dependent domains because they unify and extend traditional linear space-time models (e.g., the Kalman filter), linear and normal forecasting models (such as the auto-regression moving average method) and simple dependency models (representative of which are hidden Markov models).

⁶⁶ In the Medicine, for example, through Bayesian networks, one can represent the probabilistic relationship between diseases and symptoms and calculate the probability of the presence of any disease.

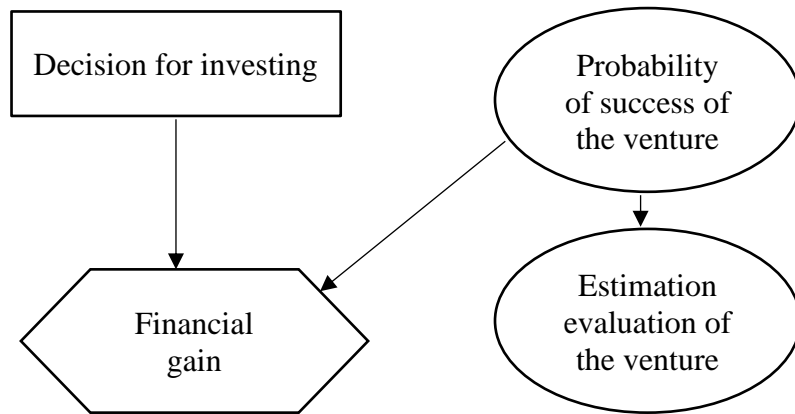


Figure 2.6. Influence diagram when planning an investment decision.

Source: author's illustration.

Compared to symbolic logic, Bayesian inference is a computationally expensive technology - the correctness of each inference must be verified by a set of conditionally independent observations. Complex cyclical solution search algorithms require the implementation of advanced statistical methods (e.g., Markov chains Monte Carlo). Bayesian networks are used in the Xbox Live gaming web service to evaluate and connect players, and in the AdSense advertising service they are used to display user-relevant websites.

Utility-based artificial intelligence systems select and plan their actions to achieve the ultimate goal using different principles from Decision theory, Decision analysis, and Information value theory: dynamic influence diagrams, Markov decision process, Game theory, and mechanism design.

The Markov decision process is a discrete time process of stochastic controlling, giving the mathematical framework for modelling decisions in situations where the results are partly random and partly under the control of the decision maker. The method is suitable for use in optimization problems that can be solved through dynamic programming and reinforcement learning.

Game theory studies mathematical models of strategic interaction between rational decision-making participant agents.

Mechanism design is a field in Economics and Game theory, through which goal-based mechanisms or incentives are created in strategic application areas where participants act rationally. The method is called reverse Game theory because the algorithm starts from the end goal and moves backwards to the beginning.

2.3. Machine Learning Algorithms

The process of programmed or non-programmed execution of a particular artificial intelligence system's task is based on the application of a particular machine learning algorithm, building the mathematical model of the problem being solved on the basis of examples of data of a variety of nature – text dictionaries, a collection of images, user records for the use of a service, etc. Typically, machine learning models are trained over a representative sample of a large population of data (training set) through a sequence of computational iterations to minimize mis execution and achieve the set goal. A potential

problem in the learning process is the creation of models that over-fit a particular training set that will not be able to work with new data or return relevant results.

Since the process of algorithmic learning to solve complex problems always costs a large time resource, the following methods are applied in artificial intelligence systems to accelerate this process:

① Introducing a degree of sufficiency of the solution found. According to this method, the learning algorithm is programmed with some pre-set error value (other than the global one) and does not look for the ideal solution to the specific problem. The principle of sufficiency reduces the number of training iterations to recognize the examples of the training set and transforms the error function from a variable into a monotonically decreasing one.

② Dynamic step management between training iterations towards optimal global error reduction.

③ Reorganization of the trainer set into recognizable classes of examples. The existence of exceptions and vaguely defined classes makes it difficult to divide the examples in the training set. The solution to this problem may be to move the exceptions to other classes or to form new classes with less variance value.

If the training set contains conflicting and unevenly distributed data, it can be reorganized by:

- Reducing the number of example classes by uniting for the purpose of uniform distribution and completeness of the set. To improve the uniformity of data distribution (it is advisable to apply a classical normal distribution of values), classes are selected that are "recognized" by the learning set. However, reducing the total number of recognizable classes also reduces the accuracy of solving the problem. Therefore, in the design process, it is necessary to reconcile the sufficiency of the number of recognized classes with the size of the trained set.
- Formation of new classes of examples comprising examples with close scatter and approximating to the reference of the relevant class.

The method of reorganization speeds up the learning process by reducing the size of the training set or by improving its quality (which is achieved after moving objects between classes and forming new classes).

④ Adaptive simplification of the training set in tasks with ever-increasing complexity. One way to reduce the complexity of the training set is to artificially approximate the output values of closely spaced examples. In this method, the output values of the examples of the simplified training set are calculated as the average of the output values of the examples from the initial sample of the training set, weighted by the function of the distance to the input values.

Adaptive simplification modifies the learning process to reduce the excess of details in the training set in the early stages of training, learn general trends and regularities in the data and ignore the errors present in the initial sample. During the training iterations, the data in the set approaches its initial values by repeating them or providing sufficient accuracy of the solution of the task.

According to the approach used, the type of input and output data and the type of the solved problems, several types of machine learning algorithms are distinguished.

1) Supervised learning algorithms build a mathematical model of a training set consisting of fully labelled training examples, presented in the form of an array or vector of matrix-arranged input data and desired outputs. By iteratively optimizing the target function, algorithms learn the functional dependence⁶⁷ between output and input and predict changes in desired outcomes when new input data is added.

2) Unsupervised learning algorithms work with training sets of unlabelled, unclassified, or uncategorized input data, identifying common similarities for their grouping or clustering (discovering potentially useful groups of input examples).

3) Semi-supervised learning algorithms combine a small number of labelled data (generally requiring costly special skills or physical experiments) with a large number of unlabelled data. In many cases, such a combination can lead to a significant improvement in the accuracy of learning.

4) Reinforcement learning algorithms train intelligent agents to perform actions with which they can achieve maximum performance in their surroundings. Because of its universality, the problem is also studied in other scientific disciplines.

5) Self-learning algorithms are applied to systems with one input (situation) and with one output (action or behaviour). The training is conducted without external reward and advice from trainers.

6) Algorithms for the learning of features in the input data preserve the input information and transform it into a form suitable for classification or forecasting. The use of such allows solving problems for reconstruction of input data with unknown distribution, for automatic design of data characteristics, for training in recognition and use of certain characteristics of data in solving a specific problem.

7) Sparse dictionary with fragmentary data learning algorithms detects features in a learning set represented as a linear combination of basic functions forming a matrix with predominantly missing values. The method is computably complex and difficult to solve approximately.

8) Anomaly detection algorithms identify errors, outliers, innovations, noise, and data exceptions.

9) Algorithms for detecting an association between variables based on associative rules identify and use a set of relational rules for a common representation of the knowledge that the artificial intelligence systems work with.

10) Federated learning algorithms decentralize the learning process to different devices by not sending user data to a central server (such a procedure for predicting search queries of mobile phone users is embedded in the Gboard application, which are not sent as separate interpellations on Google's servers).

11) Deep learning algorithms are inspired by neural doctrine and include a set of methods based on artificial neural network technology (Dechter, 1986). The architecture of deep learning tools is built from multiple cascading organized layers (Figure 2.7) of nonlinear

⁶⁷ To establish the functional dependence between output and input, the principle of 'Ockham's razor', the support vector machine approach, Bayesian network, decision trees and k-nearest neighbour methods, and computational learning theory (the principle that any wrong hypothesis will almost certainly be detected after observing a small set of examples, since it will create false predictions, and any hypothesis that consists of a sufficiently large number of correct examples will very likely be correct) are applied.

artificial neurons extracting and transforming features from unstructured, disordered, and uncategorized input data.

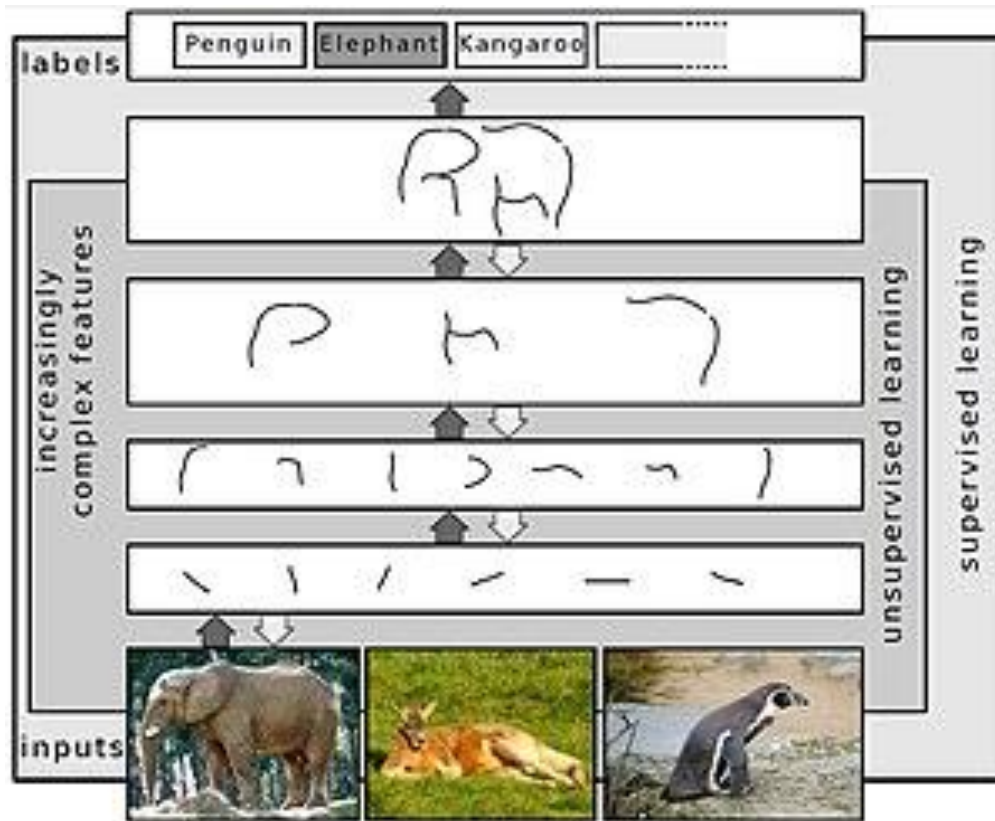


Figure 2.7. Application of the deep learning algorithm in an elephant image recognition task.

Source: Schulz, H., & Behnke, S. (2012, May 17). *Deep learning: Layer-wise learning of feature hierarchies*. *KI – Künstliche Intelligenz*, 26, 357-363.

The illustrated learning architecture mimics the multi-layered process of hierarchical extraction and presentation of data in the form of complex abstractions, carried out in the primary sensory centres in the cortex of the human brain. Each architectural layer learns supervised (in classification tasks) and/or unsupervised (in pattern analysis, for example) to transform input into some more abstract and combined representation. In the process of deep learning, artificial intelligence system independently learns how to optimally represent the individual characteristics in each layer, although sometimes it is necessary to manually adjust the number of layers to achieve the desired degree of abstraction.

Deep learning artificial intelligence systems are based on different architectures for the design of multilayer artificial neural networks (deep neural networks, deep belief networks, recurrent neural networks, and convolutional deep neural networks) and transformers. Transformers are deep learning models with a built-in mechanism for simultaneous differentiated contextual processing of all parts of successively entered input data (Figure 2.8).

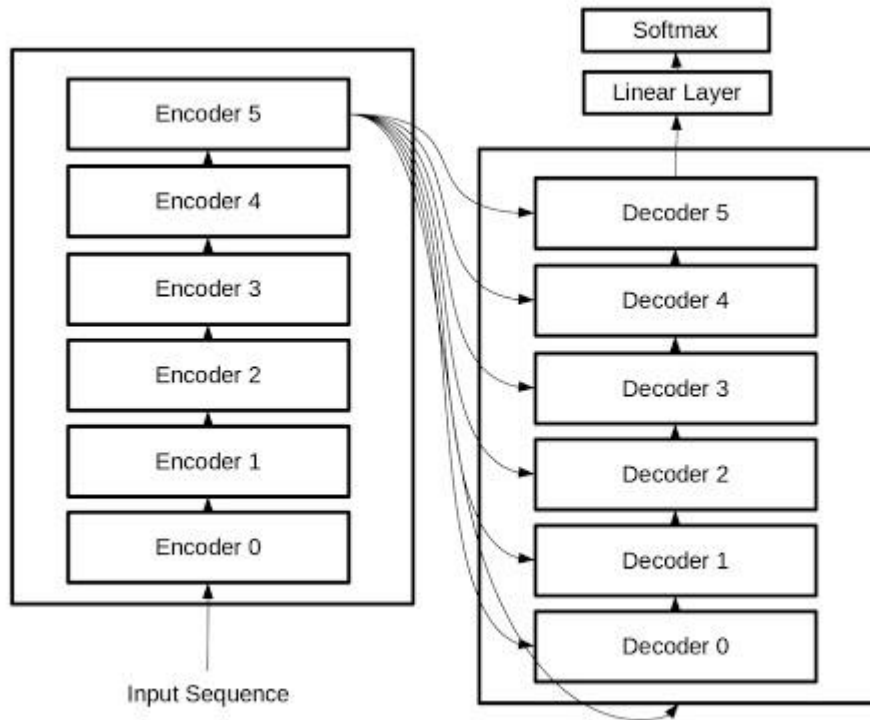


Figure 2.8. Deep learning transformer architectural layers.

Source: Hooke, K. (2020, July 23). A deep dive into the transformer architecture – the development of transformer models. <https://dzone.com/articles/a-deep-dive-into-the-transformer-architecture-the>

As can be seen from the figure above, the architecture of the transformers is encoding decoding: the encoding layers sequentially process the input data by generating binary tokens for each part of the input, with the contextual information in which each decoding layer generates an output series of data.

Transformers are trained to independently solve language modelling problems, predict subsequent text structures, answer questions, analyse sentiment in data, text summarization, document generation, semantic machine reading, machine translation, paraphrasing, recognize named entities, predict time series, and understand video data. Their training takes place in two stages: prior unsupervised training on a large set of training examples and supervised subsequent fine-tuning of unmarked examples. Pre-trained deep models such as word2vec (2013), XLNet (2014), BERT (2018), GPT-2 (2019) and GPT-3 (2020) can be adjusted to solve another problem in the field of language processing or visual perception without the need to create a new model for it.

The remarkable success of deep learning algorithms in solving intelligent problems at a level close to or superior to human performance continues to attract the interest of researchers, investors, business organizations and government structures to this sub-field of Artificial Intelligence. Artificial intelligence systems for deep learning require the availability of specialized hardware with the ability to perform matrix and vector operations with a high degree of parallelism (between 10^{14} and 10^{17} operations per second) and large training data sets.

2.4. Artificial Neural Networks

Artificial neural networks are a tool in Artificial Intelligence, developed to emulate the way the nervous system of biological entities functions. These are highly distributed dynamic artificial intelligence systems with a directional graph topology that process information from the surrounding environment by changing their state in response to a constant or impulse input signal.

Artificial neural networks are built from a set of artificial neurons that receive input data, change their internal state (are activated) according to this data, and derive input- and activation-dependent output values. The way artificial neurons' function resembles that of their biological counterpart. The mathematical model of an artificial neuron proposed by McCulloch and Pitts and shown in Figure 2.9 below is still used today as a standard in describing the architectural scheme and the actions that are implemented in this kind of processing computational elements.

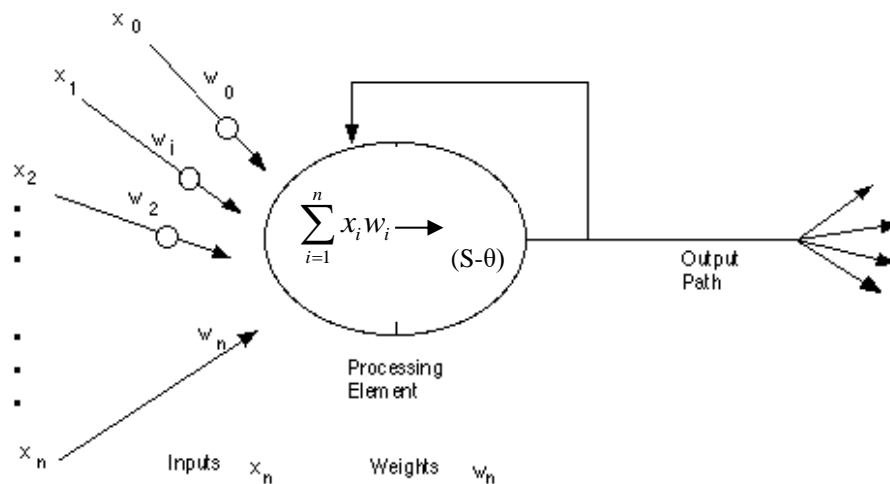


Figure 2.9. McCulloch and Pitts' mathematical model of an artificial neuron.

Source: McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. Bulletin of Mathematical Biophysics, 5, 115-133.

The input vector of values $x_0, x_1, x_2, \dots, x_n$ corresponds to the signals coming to the neuron and encompasses the set of output signals propagated from the other neural elements. Each input signal is multiplied by the corresponding link weight $w_0, w_1, w_2, \dots, w_n$, which is a positive (at activation) or negative (at snagging) scalar quantity. The weighted signals enter the summation block located in the body of the cell where their algebraic summation takes place, and the activation level S is determined. The weights of the connections, as well as the functions by which activation is calculated, can be changed by means of learning, which is regulated by some learning rule.

Other researchers who are developing models to represent the basic properties of artificial neurons and networks include Frank Rosenblatt (Rosenblatt, 1957), Fukushima (Fukushima, 1975), Grossberg (Grossberg, 1976), Kohonen (Kohonen, 1982), Hopfield (Hopfield, 1982), Anderson (Anderson, 1983), Carpenter (Carpenter & Grossberg, 1987),

Hinton (Hinton & Sejnowski, 1983) and other. The beginning of modern modelling of artificial neural networks was laid in 1982, when John Hopfield formulated the mathematical model of associative memory using Hebb's rules for programming the network and introduced a function to calculate the energy of artificial neural network as an analogue of the Lyapunov function in dynamical systems.

The process of training artificial neural networks to process information from the surrounding world is implemented in four steps:

- 1) At the beginning of the learning cycle and after establishing the sequence of switching between the learning and recall schedule⁶⁸, the neuron receives input signals through several input channels, which are initial data or output signals from the other neurons of the network.
- 2) Each input signal passes through a compound possessing a certain intensity (weight w which corresponds to the synaptic activity of the biological neuron). Each neuron is associated with a certain threshold meaning (the output from the threshold elements is established at level 0 or 1 depending on whether the sum signal of the neuron input exceeds the set threshold meaning).
- 3) In the body of the neuron through various summation functions, the weighted sum of the inputs is calculated, the threshold meaning is subtracted from it, and as a result the magnitude of the activation of the neuron (the so-called post-synaptic potential) is obtained.
- 4) The activation signal is converted using a specified activation or transfer function – hyperbolic, tangent, linear⁶⁹, sigmoid⁷⁰, sinusoidal, etc. As a result, the output signal of the neuron is obtained.

Some types of artificial neural networks necessarily require operator training, while others work independently (George & Carmichael, 2015, p. 63). Artificial neural networks can change dynamically during the learning process. The configuration of components in dynamic artificial neural networks is more complex, but the training time is shorter. The process of learning is subject to the following basic rules:

- 1) Hebbian learning rule (Hebb, 1949) describing the amplification of connections between simultaneously active neurons.
- 2) Hopfield's rule to strengthen or weaken the weights of connections between neurons according to equally active or passive input/output data.
- 3) Gradient descent learning rule to correct the delta error before applying to a neural connection.
- 4) Delta learning rule for the continuous strengthening of input neural connections in order to reduce the difference between the desired and the actual outcome of the neuron output.

⁶⁸ Recall is the process of entering input into the trained neuron and receiving a response, while in the learning process the neuron automatically extracts functional dependencies from the input data and remembers them as weights and connection structure for subsequent use.

⁶⁹ In this activation function, the output activity is proportional to the total weighted input of the neuron.

⁷⁰ In sigmoid dependence, the output varies continuously with the input signal, following the nonlinear variations of the input.

- 5) Kohonen's learning law of contested competition between neurons in the process of learning and modifying the weights of their connections (the neuron with the highest volume of input information is proclaimed a "winner" and has the ability to silence other neurons).

Like any computational network, the neural network has input elements (taking the meanings of the variables of the surrounding world) and output elements (creating predictions or control signals). Between the beginning and end of the network can be located various intermediate (hidden) neurons performing specific intra-network functions. The successful coupling in an artificial structure of input, output and hidden neurons depends on the ordering of neurons, the choice of activation function and connections' weights (the latter are a major carrier of knowledge, implicitly represented in the general architecture and interaction between neurons).

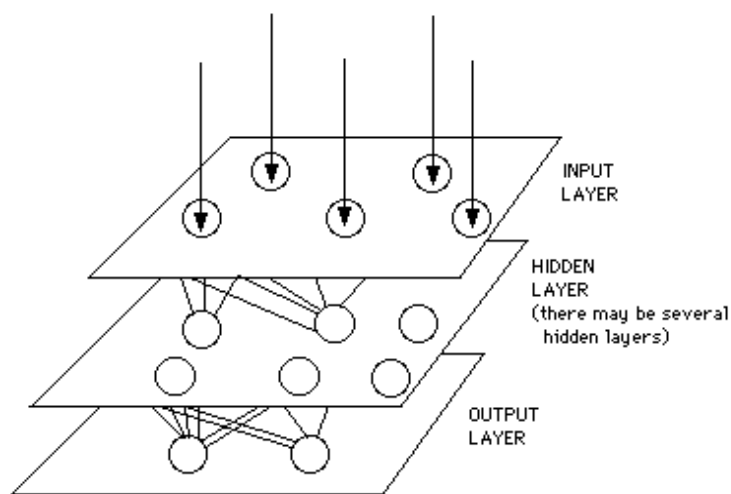


Figure 2.10. Architectural scheme of a feedforward two-layer artificial neural network composed of input, hidden and output layers of artificial neurons.

Source: author's illustration.

The biological neural network is distributed in a three-dimensional world of microscopic elements that is difficult to realize in artificial neural networks, which are essentially two-dimensional devices built of clustered in separate layers of neurons. The number of layers and the number of neurons in each layer determines the architectural scheme of the network and its characteristics (Figure 2.10).

Different layers perform different types of transformations on the input data. Signals move from the first input layer to the last output layer of the network, often passing several times through the different layers. When this movement is only in the input-output direction, the artificial neural network is feedforward, and when in addition to straight there is a reverse information flow, it is recurrent.

The existing variety of artificial neural networks is most often classified by the characteristic as topology, ways of solving problems and purpose.

According to the type of topology, which can change in order to more fully match the problem solved, single-layer and multilayer artificial neural networks are distinguished. In single-layer artificial neural networks, each neuron can play the role of both input and output elements, connecting with all other neurons (full-connected artificial neural networks), or only

with the two closest neurons along a horizontal and vertical border (regulatory⁷¹ artificial neural networks). In multilayer networks (Figure 2.9) there are two external layers (input and output) and one or more internal layers performing computational data processing. The connections between neurons in single-layer and multi-layer networks are feedforward, crossbar, recurrent and lateral.

According to the way of solving the problem (formal, informal, or mixed) artificial neural networks are divided into forming artificial neural networks, artificial neural networks with a forming matrix of connection, learning artificial neural networks⁷² and combined artificial neural networks (multilayer networks in which each layer is represented by a different topology and trained by a certain algorithm).

According to their purpose, artificial neural networks are classified into five categories. The learning algorithms and application areas of each of the categories are summarised in Table 2.1. A more detailed examination and illustration through graphical schemes of the characteristics of almost each artificial neural network in this classification is made after the tabular systematization.

Table 2.1. Classification of artificial neural networks according to their purpose.

Source: author's systematization.

Artificial neural network category	Learning algorithm	Application
Artificial neural networks for prediction	<ul style="list-style-type: none"> • Back-propagation • Delta bar delta • Extended delta bar delta • Directed random search • Higher-order neural networks • Self-organizing map into back-propagation 	Selecting the best goods on the market, synthesizing speech from text, processing images, predicting the weather, disease risks, and so on.
Artificial neural networks for classification	<ul style="list-style-type: none"> • Learning vector quantization • Counter-propagation • Probabilistic neural networks 	Establishment of classifications, image and video segmentation, recommendation scoring systems and natural language processing

⁷¹ The regularity of the network is defined by the strict definiteness of the number of connections of each neuron, which depend on its location in the interior, node or boundary of the one-dimensional network. Connecting boundary neurons from two parallel boundaries transforms the planar structure into a two-dimensional cylinder. If it is necessary to connect the boundary neurons of the cylindrical structure with two other parallel boundaries, the two-dimensional cylinder is transformed into a multidimensional tor. The presence of a large number of neural connections makes multidimensional multilayer artificial neural networks difficult for apparatus-program applicational implementation.

⁷² In the process of training the learner artificial neural networks some of its parameters (usually the coefficients of the synaptic connection or the topology) change automatically. The training time is still large, which is the main disadvantage of this type of applications. Therefore, the choice of training algorithm is of fundamental importance.

Artificial neural networks for data association	<ul style="list-style-type: none"> • Hopfield network • Boltzmann machine • Hamming network • Bidirectional associative memory • Spatio-Temporal Pattern Recognition 	Recognition of classification errors (for example, errors in the symbols of scanned text), recognition of repetitive audio signals, etc.
Artificial neural networks for data conceptualisation	<ul style="list-style-type: none"> • Adaptive resonance network • Self-organizing map 	Analysing information (for example, to extract names of people who would buy a particular item according to certain criteria)
Artificial neural networks for data filtering	<ul style="list-style-type: none"> • Recirculation • Kalman Filter 	Correction of input signal, filtering objects from static and dynamic application domains, solving matrix equations, facial recognition, signal processing and pattern recognition

The feedforward back-propagation artificial neural networks consist of an input, output and at least one hidden layer fully connected to each other (Figure 2.11). back-propagation algorithm (Rumelhart, Hinton, & Williams, 1986b; Parker, 1987) is the most popular, efficient and easy to implement model for the construction of complex multilayer artificial neural networks, applicable to different categories of networks with radically different topologies, training methods and a wide range of tasks in which a nonlinear solution to poorly formalized problems is sought.

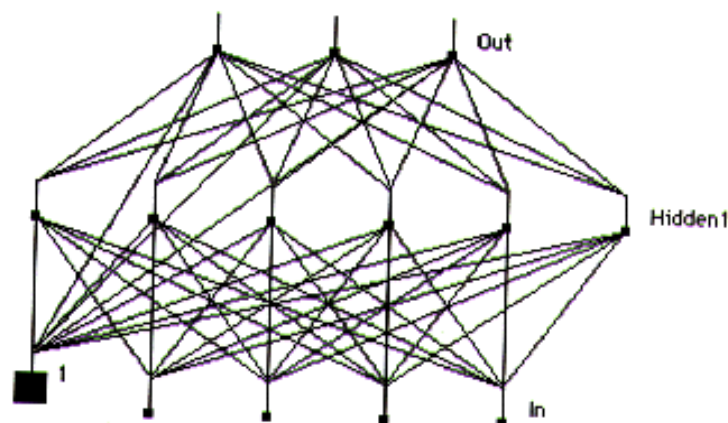


Figure 2.11. Topology of feedforward back-propagation artificial neural network.
 Source: NeuralWare, Inc. (1991). *Neural computing: Neural-works Professional II/Plus ANN development software*. Pittsburg: NeuralWare, Inc.

The learning process in feedforward artificial neural networks usually uses some variant of the Delta rule, which begins by calculating the difference between actual and desired outcomes. Using this error, the weights of the connections increase in proportion to the error time, which is a scaling factor for the overall accuracy. To accomplish this, the inputs, outputs and desired output must be represented at each individual node of the artificial neural network in the same artificial neuron. The complexity of this learning algorithm is that the network can hardly determine which input contributed the most to the incorrect output and how this element changed to correct the error (inactive nodes do not affect the error and the weights of their connections do not change).

To solve this problem, the training input data is entered at the input layer of the network, and the desired results are compared with the achieved ones at the output layer. During the learning process, the structure of the network is modified, and the output of each neuron is calculated layer by layer. The difference between the output of the final layer and the desired output is returned changed to the previous layer(s) by a derivative of the transfer function and adjusted using the Delta rule weights of the connections. This process continues for each preceding layer until the input layer of the network is reached.

The back-propagation algorithm is based on the 'gradient descent' method⁷³, which minimizes errors in the artificial neural network predictions during the process of changing the connection weights of each artificial neuron. In this approach, a standard convergence speed is used in each individual layer, and the momentum rate is set in total for the entire network. Some back-propagation algorithms allow the convergence speed to gradually decrease as large amounts of learning sets pass through the network.

There are different variations of the back-propagation artificial neural networks learning rules. Various functions can be used to calculate the error (momentum error, cumulative reverse back-propagation), transfer functions and even a modified transfer function derivative method.

The feedforward back-propagation architecture has several limitations: it requires continuous supervised training using many input-output examples, there is no guarantee of reaching an acceptable solution by the artificial neural network, the training gets stuck when an error is found less significant than the theoretically smallest possible one (that's why in many artificial intelligence systems a computational time is added, after which such errors are considered temporary and ignored).

Typical tasks in which the feedforward back-propagation algorithm is applied are the synthesis of speech from text, the evaluation of bank loans, image processing, the representation of knowledge, the simultaneous tracking of several targets, etc.

The delta bar delta artificial neural networks are networks with continuous input signals and continuous activation functions, in which the tuning of the weight of each neuron is based on the following formula (Jacobs, 1988):

$$Dw = \eta (d - \text{Out}) x, \text{ where:}$$

⁷³ 'Gradient descent' is a learning method based on setting an evaluation function that measures the error in the operation of the artificial neural network as a differentiable function of the weights. Each possible combination of neuron weights specifies a point on the error surface that can have a positive or negative slope. When the absolute value of the error becomes 0 or as small as possible, the training goal is reached (Atanasova, 2005, p. 164).

Δw – change in the weight of the neuron

η - training rate coefficient

d – desired output

Out – achieved output (when d and Out coincide, it is not necessary to change the weights of the connections)

x – input meaning of the neuron

The delta bar delta artificial neural networks use a learning method where each weight has its own self-adapting coefficient. A characteristic of this method is that it does not use a momentum factor, but all other operations in the network (e.g., feedforward recall) are identical to the back-propagation architecture. The delta rule is a heuristic learning approach, which means that recent error values can be used to assume the magnitude of future ones. Knowing the probable errors allows the system to take intelligent action to correct the weights of connections between neurons. However, this process is complex and there is no empirical evidence that each weight can exert a different effect on the overall error.

The rules directly applied to this algorithm are understandable and easy to implement. Each weight has its own learning rate, which changes based on the information about the current error found by the standard back-propagation method. When the weight of the connection changes, if the local error has the same sign for several consecutive steps, the learning rate of this connection increases linearly. When the local error often changes its sign, the learning rate geometrically decreases (the latter ensures that the learning rate of the connections are always positive and can be reduced in artificial neural network' areas where the change in error is large). By assigning a different learning rate to each weight of a connection in the network, the need to search for the minimum of the error by the steep descent method (in the direction of the negative gradient) is eliminated.

Extended delta bar delta artificial neural networks are networks with exponential learning rate delay, added momentum coefficient and fixed upper limit of learning rate (Minai & Williams, 1990).

This algorithm uses the sign of the current error to assess whether the learning rate should be increased or decreased. The reduction occurs in a way identical to that applied in delta bar delta networks. In the inverse process, the learning rate and the momentum coefficient are modified to become exponentially decreasing functions of the magnitude of the weighted gradient components. Thus, larger increases are applied to areas with small slopes or curvature. But this is only a partial solution to the jump problem characteristic of the delta bar delta. To prevent jumps and oscillations in the weights of neural connections, upper limits are placed on individual connection learning rates and momentum rates.

In the algorithm of Minai and Williams a memory with the possibility of recovery is built. After each epoch of presentation of the training set through artificial neural network an accumulated error is evaluated. If it is less than the previous minimum error, the weights of connections between neurons are saved to memory as current best values. The recovery phase is controlled by a tolerance parameter - if the current error exceeds the minimum previous error modified by the tolerance parameter, all values of the connection weights regress stochastically in the direction the best set of weights stored in memory. Furthermore, the learning and momentum rates are reduced to initiate a process of recovery of artificial neural network.

Directed random search artificial neural networks have a standard feedforward recall structure in which the weights of connections between neurons are set randomly by directed search. In order to achieve optimal self-tuning, a component is added to the random step which approximates as much as possible the weights of connections between neurons to the optimal ones. The area of best functioning of the connection weights of each individual neuron is established within wide borders, within which the directed random search algorithm generates a set of initial randomly distributed connection weights for each of the neurons.

Directed random search artificial neural networks have four components - random step, reversal step, directed component and self-adjusting variance. The directed random search paradigm is characterized by greater speed of operation and easy applicability in solving clearly defined and relatively simple problems. In contrast to computation-based techniques (i.e., the delta rule and its variations), the training loops in this algorithm are much shorter as the likely errors in the intermediate neurons are not calculated, but only the error of the output of the network.

Directed random search is automatic, requires little or no user interaction and is easily applicable to smaller networks with a small number of neural nodes. With artificial neural networks with more than 200 connection weights, a relatively long training time is required, which does not always guarantee obtaining an acceptable solution.

Higher-order neural networks or functional-link networks (Pao, 1989) have a standard feedforward back-propagation architecture, extended by including additional nodes in the input layer. A characteristic of these is that the input data are transformed by basic mathematical functions into higher-order functions or into functionally related transformations such as squares, cubes, or sines. This algorithm can drastically improve the learning rates in some artificial intelligence systems because the higher-order functions are applicable to any feedforward back-propagation artificial neural networks using the delta bar delta or the extended delta bar delta.

Adding of additional input nodes in the artificial neural network is done either by adding cross-product of the input terms (tensor model,) or by functionally expanding the base data. The result of both ways is the creation of a network with the possibilities for a more comprehensive presentation of the input data. A higher order of input data can make the artificial neural network easier to train because joint or expanding activation functions are directly implemented in the model. In some cases, the need for a hidden layer is even eliminated.

The limitations of functionally-link artificial neural networks are that in order to use the transformed input data many more input nodes must be processed. As the order of the calculations carried out by the artificial intelligence systems increases, this problem becomes worse. Therefore, the input data should not be increased more than necessary.

Back-propagation self-organizing map artificial neural networks are hybrid networks having a self-organizing map for conceptual data partitioning prior to their use by the familiar back-propagation algorithm. The self-organizing map helps to visualize topologies and hierarchical structures of higher-order input spaces prior to their input into the feedforward back-propagation artificial neural network. The change in input occurs along the lines of the automatic functionally-link input structure. The self-organizing map is trained in an

unsupervised way, and the rest of the network is trained supervised (Kohonen, Self-organizing maps, 1997).

Deep neural networks are feedforward artificial neural networks with multiple layers hidden between the input and output layer. These types of artificial intelligence systems find the right mathematical operations to convert input data to output, regardless of whether the relationship between neurons is linear or nonlinear (Hinton, 2007). On each of the layers, the probability of the different possible outputs is calculated. The user can view the results and choose to display only the responses with a probability percentage above a threshold set by him.

Deep artificial neural networks can model complex non-linear relationships. Their architecture creates compositional models in which objects are expressed in the form of multilayer compositions of primitives (basic or embedded data types). In the additional layers of the network, characteristics of the lower layers can be combined and complex data with fewer processing units can be modelled.

Learning vector quantization artificial neural networks (Kohonen, Learning vector quantization, 1988) are based on the so-called Kohonen layer, whose function in the architecture of artificial neural network is the sorting of similar objects into corresponding classes (therefore the algorithm is particularly suitable for classification and image segmentation tasks). The topology of artificial neural networks of this kind contains an input layer, a single Kohonen layer and an output layer (Figure 2.12).

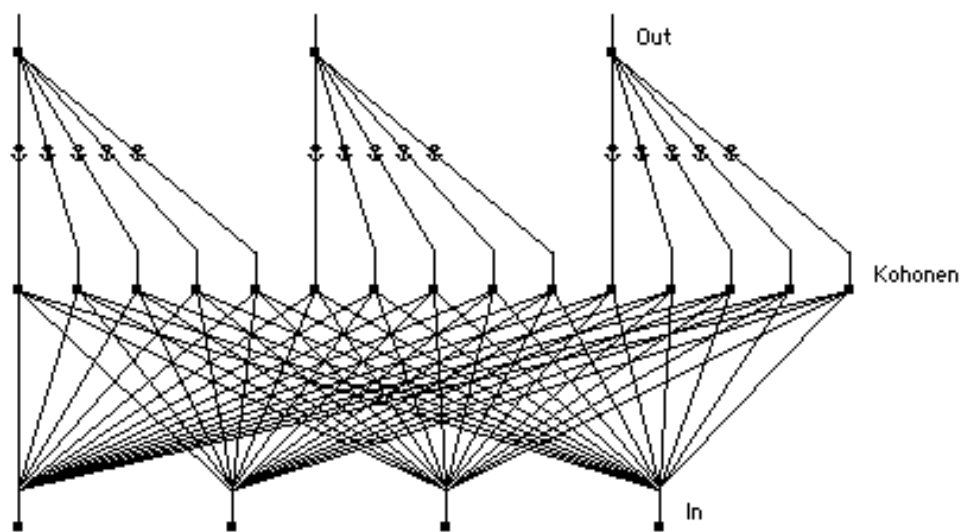


Figure 2.12. Topology of learning vector quantization artificial neural network.

Source: NeuralWare, Inc. (1991). Neural computing: Neural-works Professional II/Plus ANN development software. Pittsburg: NeuralWare, Inc.

There are as many neurons in the output layer as there are different classes of objects. In the Kohonen layer there is a uniform set of neurons to process each class (in the example of five), the number of which depends on the complexity of the input/output interactions. The training of the data set and the relational classification of the objects in it is supervised in the Kohonen layer according to rules different from those in the back-propagation artificial neural

networks. In order to optimize the training and recall activities of the network, the input layer is constructed with one neuron for each individual input parameter or higher order input structures are used.

Vector quantification learning artificial neural networks classify the input data while retaining the internal topology of the training set. The training algorithm uses Kohonen's layer to calculate the distance from the learning vector to each neuron. The nearest one is declared the 'sole winner' of the layer and determines the class to which the input vector belongs. If the winning neuron is of the class of the learning vector, the weight of its connection is sent to the vector, and if it is not, it is removed from the learning vector. All neurons assigned to a class cluster within the layer.

In complex classification problems with similar objects or input vectors, the vector quantification learning algorithm requires a large number of neurons in the Kohonen layer. This problem is overcome by better selection or use of higher order representations for the input parameters.

Another problem is the 'propensity' of some neurons to win too often. This can be avoided by:

- Adding a conscience mechanism – a distance bias that is proportional to the difference between the frequency of victories of a given neuron and the average frequency of victories of all neurons. After each passage of the training set through the artificial neural network the distance bias step is adjusted in the direction of reduction.
- Application of a boundary adjustment algorithm, which is suitable in cases where the winning neuron is taken to the wrong class, and the second best one is placed in the right one. The winning neuron moves farther away from the learning vector, and its place is taken by its runner-up.
- Using an attraction function at different points while training the artificial neural network.

Counter-propagation artificial neural networks (Hecht-Nielsen, 1987) combine the unsupervised Kohonen layer with a learning output layer, minimizing the number of neurons and the network training time. The topology of these networks (Figure 2.13) consists of an input layer, a buffer layer (used to normalize data before processing), a self-organizing Kohonen layer (acts as an adaptable lookup table finding the closest fit of the input data and inferring their equivalent mapping) and an output layer (uses the delta rule to change the input weights of the connections).

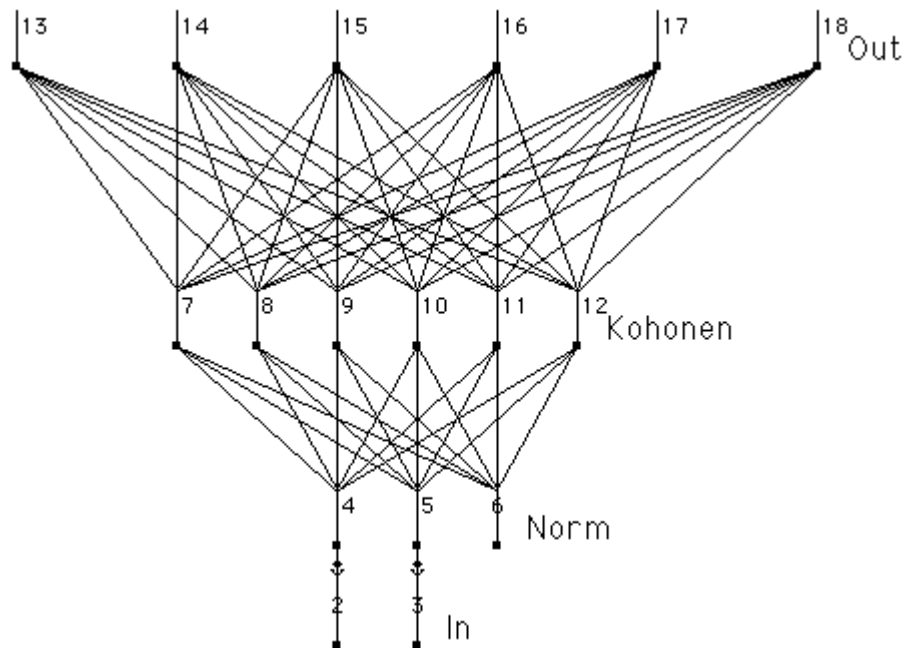


Figure 2.13. Topology of counter-propagation artificial neural network.
Source: NeuralWare, Inc. (1991). Neural computing: Neural-works Professional II/Plus ANN development software. Pittsburg: NeuralWare, Inc.

The size of the input layer depends on the number of individual parameters defining the problem solved: if it is too small, the artificial neural network cannot generate enough generalizations, and if it is too large - the processing time will be very long. The normalization layer contains one neuron for each input example plus another balancing neuron. Thanks to normalization, the Kohonen layer classifies the examples of the learning set into self-organized zones of classes. The output layer uses the delta rule to adjust the weights of the connections entering it in the direction of achieving the desired result.

A possible problem that may arise with this type of artificial neural network is the unsupervised classification of objects in the Kohonen layer (a condition is created for a neuron of this layer to learn to operate with two or more input examples belonging to different classes). To avoid the problem, neurons in the Kohonen layer can be pre-taught to operate a certain class of objects.

Probabilistic artificial neural networks (Specht, 1988; Specht, Probabilistic neural networks, 1990) solve supervised example classification problems using Bayesian classifiers and Parzen estimators (Parzen, 1962). The topology of these networks (Fig. 2.14) consists of an input layer, a normalization layer, a pattern layer (organizes the training set so that each input vector is represented by a separate neuron), a summation layer (contains as many neurons as classes need to be recognized), and an output layer.

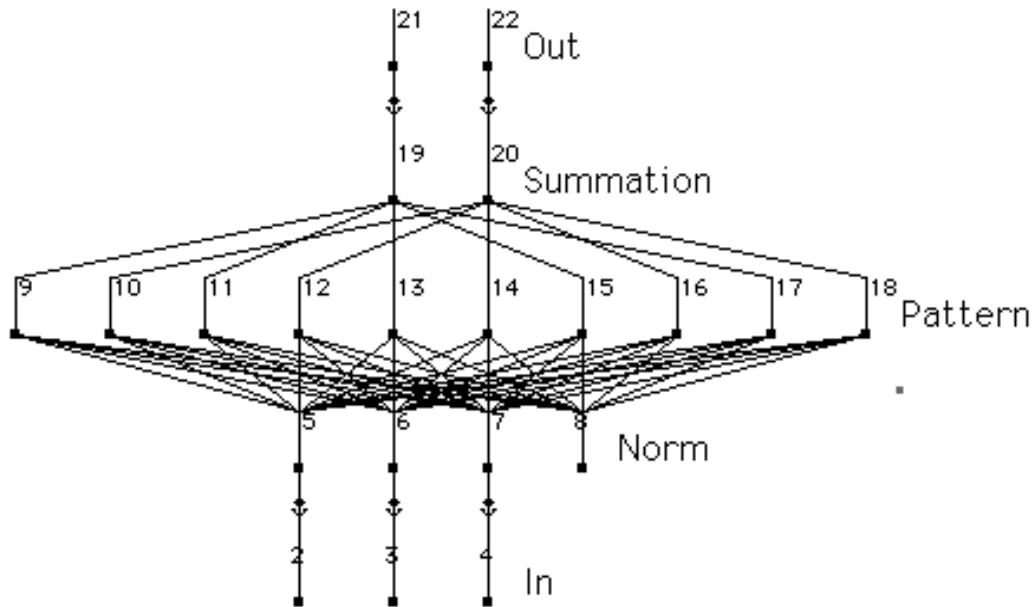


Figure 2.14. Topology of probabilistic artificial neural network.

Source: NeuralWare, Inc. (1991). Neural computing: Neural-works Professional II/Plus ANN development software. Pittsburg: NeuralWare, Inc.

The pattern layer represents a neural implementation of a version of a Bayes classifier, where the class dependent probability density functions are approximated using a Parzen estimator. It has one neuron for each input vector of the training set, which is trained only once. The model layer works on a competitive principle – only the most accurate correspondence of the input vector with the desired result can generate a classification class. Training examples can be combined or weighted with the magnitude of a previous probability of assignment to each category to determine the most likely class for a given input vector. If this relative frequency is unknown, all categories can be assumed to be equally probable, and class determination will be based solely on the proximity of the different input vector to the distribution function for a given class. The classification process is refined by including Parzen estimators estimating the frequency of occurrence of the training examples.

Training in probabilistic artificial neural networks is relatively easy. If the difference between classes varies and at the same time is quite similar for certain application areas, the pattern layer can become quite large. The main advantage of this category of networks is the mathematical foundation on which their functioning is based.

Deep Belief Networks (DBNs) represent generative graphical models composed of multiple layers of latent variables (hidden processing neurons) with connections between layers but not between neurons within each layer. When trained in unsupervised way on a data set, this type of network can learn to reconstruct input data probabilistically so that each layer detects a particular classification characteristic in the input examples. After this step, the training can continue supervised.

Convolutional deep neural networks are fully connected networks in which the mathematical operation of convolution is used (Fukushima, 1980). The architectural layers of neurons in this category of networks (Figure 2.15) realize four main operations: convolution, nonlinearity (ReLU), pooling and full connecting.

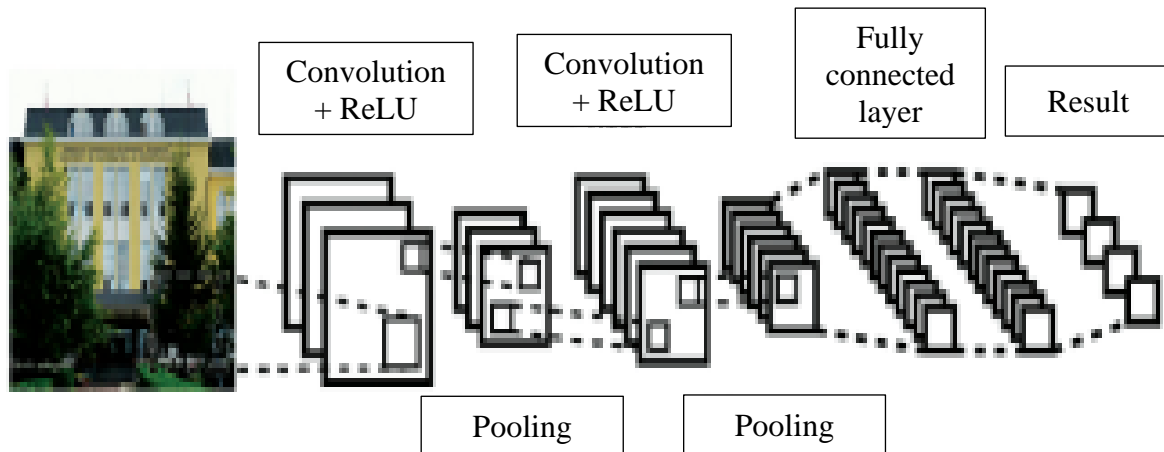


Figure 2.15. Example architecture of convolutional deep neural network.

Source: Hristov, A., Nisheva, M., & Dimov, D. (2018). *An introduction into convolutional neural networks. Automatica and Informatics (1)*, p. 28.

Most deep learning artificial intelligence systems are based on convolutional deep neural networks, which are commonly used in tasks such as image and video recognition, image classification, medical image analysis, recommendation scoring systems, and natural language processing.

The crossbar associative artificial neural network of Hopfield is constructed of three layers with the same number of neurons in each layer (Figure 2.16). Each neuron in the Hopfield layer is directly connected to the outputs of all neurons in the input layer by variable connection weights, is recurrently connected to the neurons in the output layer, and changes its state in a certain number of learning cycles until the artificial neural network reaches a steady state. For training the network in the direction of a more precise association of the classified classes, a sigmoid transfer function is applied.

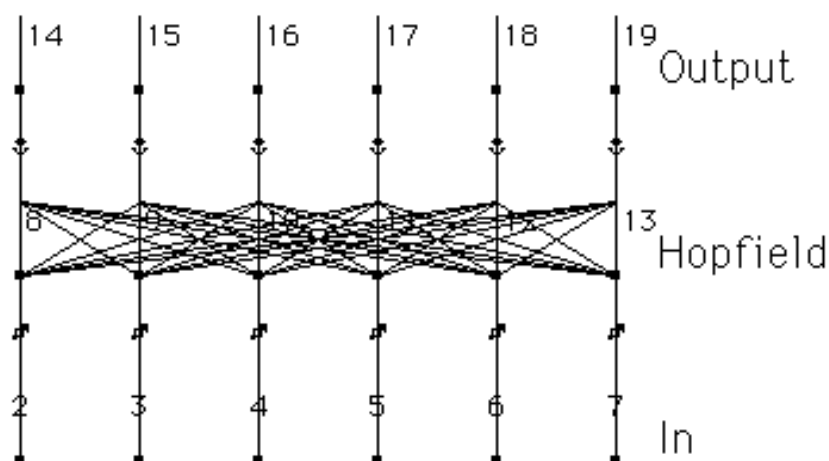


Figure 2.16. Topology of the Hopfield artificial neural network.

Source: NeuralWare, Inc. (1991). *Neural computing: Neural-works Professional II/Plus ANN development software*. Pittsburg: NeuralWare, Inc.

Training of Hopfield networks requires simultaneous use of training patterns at the input and output of the artificial neural network. The recursive nature of the Hopfield layer corrects the weights of all connections, applying the eponymous rule and transfer function with the threshold mechanism involved (since matched input/output pairs are rarely obtained the network will not be trained properly).

Hopfield's artificial neural networks has two major drawbacks: the number of pattern classes that can be stored and properly activated is approximately 15% of the number of neurons in the Hopfield layer; the Hopfield layer may become unstable if the patterns used are too similar.

The Boltzmann machine uses a technique to determine the original pattern more accurately, minimizing the total energy for the space of states of the solution sought by the technique of simulated annealing (Ackley, Hinton, & Sejnowski, 1985). According to this concept, the learning process of the network should be periodically cooled to lower the noise parameter of each neuron and the state of the network to calm down. If this does not happen, every passage of the training set through the artificial neural network will move away from reaching an optimal solution.

Once the Boltzmann machine fully associates the training set, it can also work with incomplete classes of patterns that should not exceed 15% of all neurons in the network.

The Hamming artificial neural network is an extension of the Hopfield network whereby a maximum likelihood classifier is added at the end of each input node (Lippmann, 1987). The Hamming network has three layers – input layer, category layer and output layer. In this learning algorithm, the training set passes through the network under the supervision of a teacher only once. The requested training pattern is set at the input layer, and its class is created at the output layer. The output contains only the output category to which the input vector belongs.

The first set to be fed to the category layer is the connection weights. Neurons at the category layer output generate matching results that are calculated as the difference between the number of input nodes and the Hamming distance to the examples of input vectors. These results range from zero to the total number of input neurons and are highest for those input vectors that best fit the patterns learned. The connection weights in the recursive category layer are trained in the same way as in the Hopfield network. In a normal feedforward operation, the input vector is applied to the input layer, where it is presented for as long as it is necessary to establish the matching results in the lower level of a subnet in the category layer. This triggers an introduction of the Hopfield function into the category layer and allows a part of the artificial neural network to detect the nearest class to which the input vector belongs.

The Hamming artificial neural network has several advantages over the Hopfield network. It implements a classifier of the optimal minimum error when entering random and independent errors at the level of data bits. The Hamming network needs fewer neurons because the category layer requires only one neuron per category, not for each input node. This network does not create misclassifications, it is faster and more precise.

Bi-directional associative memory artificial neural networks are used to train a set of paired patterns represented as bipolar vectors (Kosko, 1988). In the topology of this category of networks (Figure 2.17) the number of neurons in the input and output layer is equal. The

two middle layers are composed of two fully interconnected bi-directional associative memories (BAM) to represent the values of two input vectors of equal length.

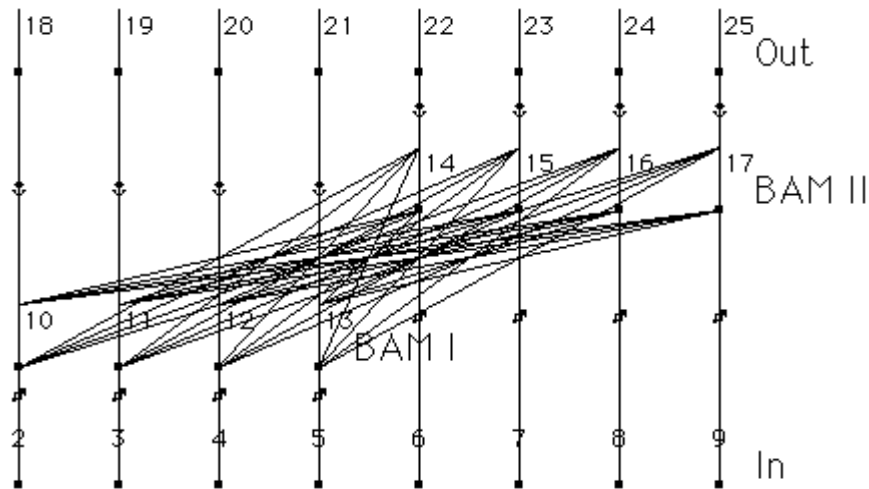


Figure 2.17. Topology of bi-directional associative memory artificial neural network.
Source: NeuralWare, Inc. (1991). Neural computing: Neural-works Professional II/Plus ANN development software. Pittsburg: NeuralWare, Inc.

The middle layers are designed to store associated pairs of vectors. When a noisy pattern of a vector is applied at the input, the middle layers begin to oscillate until a steady equilibrium state is achieved. This state, indicating that the network has not yet been trained, corresponds to the closest learned association, and leads to the generation of the original training pattern at the output. Like the Hopfield network, the bi-directional associative memory network may not find the requested training pattern if an additional training set is used as an unknown input vector.

The spatio-temporal pattern recognition artificial neural networks memorise specific patterns, which are then used as a basis for classifying input repetitive signals. The architectural scheme of these networks contains a number of tuning parameters to detect time-varying signals, and a common threshold element is attached to each neuron normalizing the overall performance of the network.

The spatio-temporal pattern recognition artificial neural networks training algorithm uses a variant of Kohonen's rule, and a time-varying component of the learning function called attack function. The main application of this category of networks is in the tasks of recognizing repetitive audio signals where, even in the presence of large differences in the periodicity of input signals, the slow attenuation of the attacking function keeps the artificial neural network functional.

In recurrent neural networks, data processing can also be done in the reverse direction (from the farther to the nearer neurons). Node's connections in recurrent networks form a directed graph with a temporal sequence, allowing temporal dynamic behaviour to be expressed and making them applicable to tasks related to language modelling, unsegmented handwriting recognition, and unsegmented speech recognition. Currently, a deep variant called LSTM - networks with long short-term memory - is widely used (Hochreiter & Schmidhuber, 1997).

Adaptive resonance artificial neural networks create categories for input data examples based on adaptive resonance theory⁷⁴. The topology of this category of networks is similar to the biological neural brain structure - at the core of the network stand two closely interconnected resonant layers of neurons located between an input and an output layer. An unsupervised learning function is used to analyse the essential input examples and to detect the possible characteristics or classification patterns in the input data vector.

Although adaptive resonance artificial neural networks are commonly used for biological modelling, they also have engineering applications. Their main limitation is the sensitivity to noise - even a small, weak noise of the input vector confuses the possibilities of matching examples from the training set.

In artificial neural networks with self-organizing maps, the input data is projected onto a two-dimensional Kohonen layer, which preserves the order of their input, compacts scarce data, and expands dense data. Input vectors with close values are mapped onto those neurons located close together in the two-dimensional layer, which represent most accurately the characteristics or examples of input data, forming their two-dimensional neural map. The self-organizing map (Kohonen, Self-organized formation of topologically correct feature maps, 1982) is mainly used for visualization of topologies and hierarchical structures of higher-order dimensional input spaces and for creating area-filled curves in the two-dimensional Kohonen layer. Typically, a self-organizing map is trained unsupervised, but if its topology is combined with other neural layers for prediction or categorization, the artificial neural network is initially trained unsupervised and then switched to supervised mode.

In Figure 2.18 is an illustrated example of an artificial neural network with a self-organizing map in which the transition layer is entirely connected to a two-dimensional Kohonen layer, and the output layer categorizes the input vector of data into three classes. The output layer is usually trained by applying the delta rule and functions similarly to the counter-propagation paradigm. The training algorithm measures the Euclidean distance between the weights of all neurons in the Kohonen layer and the input initial values. Only the shortest distance neuron for a given input vector passes to the output layer, all others are reset. The magnitude of the non-nulled neuron which is essentially closest to the value of the input example is represented in the two-dimensional Kohonen map.

⁷⁴ The theory of the cognitive process of information processing in humans was created in 1976 as an attempt to explain how the human brain autonomously learns to understand, categorize, recognize and predict objects and events in a changing world (Grossberg, 1976).

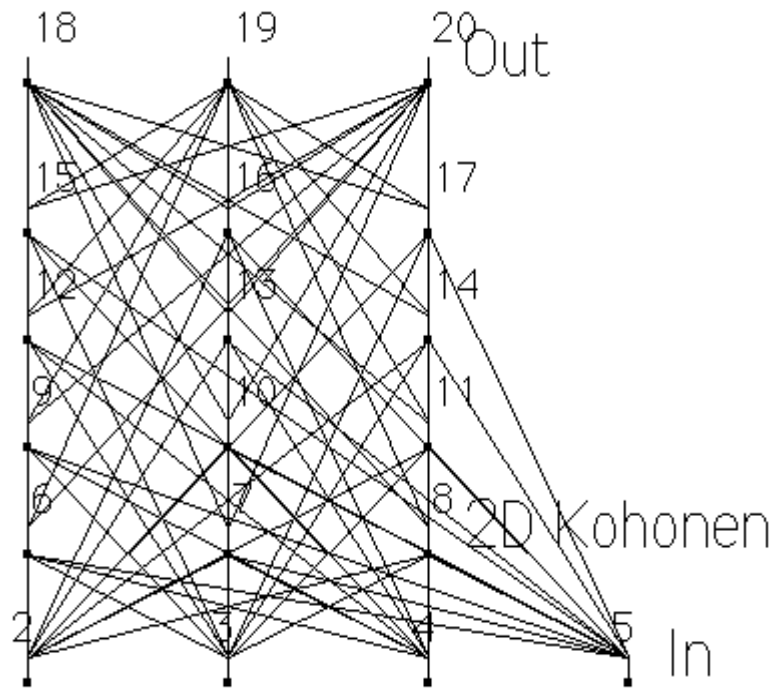


Figure 2.18. Topology of self-organizing map artificial neural network.

Source: *NeuralWare, Inc. (1991). Neural computing: Neural-works Professional II/Plus ANN development software. Pittsburg: NeuralWare, Inc.*

In the algorithm of self-organizing maps, a problem arises when one neuron must present too many input data. This is solved by embedding a conscience mechanism in the training function. Input spaces with sparse data are presented compactly in the Kohonen map, while those with high density are ‘expanded’ to achieve a more precise separation. Thus, the Kohonen layer ‘mimics’ the way of representing knowledge in biological systems.

In the recirculation artificial neural networks, data is processed in only one direction, and training is carried out unsupervised using only local knowledge generated on the basis of the state of neurons and the input values of the connections to be adapted (Hinton & McClelland, 1988). In the topology of this category of networks (Figure 2.19) there is the same number of inputs and outputs and two intermediate layers (visible and hidden). The visible layer data are compressively represented in the hidden layer by as few neurons as possible. The two interlayers are fully and bidirectionally connected, both with each other and with a bias element. Connections between neurons in these layers have variable weights and are trained similarly to other connection weights in the network.

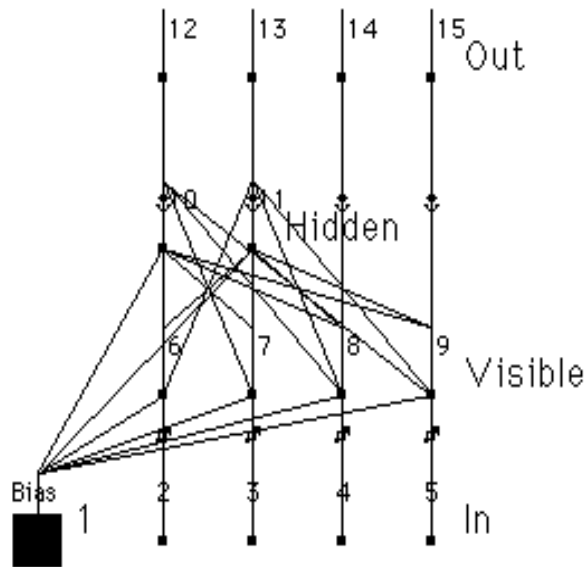


Fig. 2.19. Topology of recirculation artificial neural network.

Source: NeuralWare, Inc. (1991). Neural computing: Neural-works Professional II/Plus ANN development software. Pittsburg: NeuralWare, Inc.

The learning process of recirculation artificial neural networks is similar to that of networks with bi-directional associative memory - the input data is presented in the visible layer and transmitted to the hidden one, which returns the result back to the visible and transfers it to the output layer. As a consequence of this second pass through the hidden layer, a circulation of the input data through the structure of the network is obtained and the network is trained.

During network training, an encoded version of the input vector is initially fed to the output of the hidden layer. In its next passage through the network, it is reconstructed at the output of the visible layer to its original appearance. The purpose of the learning algorithm is to reduce the error between the reconstructed and the initial vector, the error when reconstructing the hidden layer and the differences in the output data between their first and last passage through the hidden layer. Recirculation artificial neural networks are successfully applied in matrix equation solving problems and facial recognition (Bryliuk & Starovoitov, 2001).

Filtering artificial neural networks using Kalman's algorithm apply a learning method in which each neural weight is updated according to a second-order derived value matrix generated by multiple training examples. During training, the weights of connections between neurons alternately change, and the matrix is preserved and changed.

Kalman's filtering algorithm (Kalman, 1960), used to formulate the state space of linear dynamical systems, is applied to solving problems for the linear optimal filtering of objects from movable and stationary domains. The solutions found are recursive in nature, as each updated space estimate is calculated based on the previous estimate and the new inputs. Since there is a need to store only the previous estimates and not all the observed data, Kalman filtering works much more efficiently than performing direct calculations with the observed aggregate at each step in the screening process.

Singhal and Wu (Singhal & Wu, 1989) extended Kalman's filtering algorithm, creating a sequential second order learning method for static multilayer perceptron networks that proved to be significantly more efficient in terms of training epochs than the backpropagation algorithm in problem solving by pattern series recognition. Their development became the basis for the development of many methods for artificial neural networks training with feedforward error propagation and their application in solving problems related to control, signal processing and pattern recognition.

The decoupled extended Kalman algorithm (Puskorius & Feldkamp, 1991) is an artificial neural network learning algorithm creating and maintaining secondary information about the weights of neural connections that belong to mutually exclusive groups of input data. Puskorius and Feldkamp considered that the grouping of training examples should be done according to the criterion 'network node' and developed a procedure for separating the interactions between the weights of input data connections by network nodes. Kalman's nodal decoupled extended algorithm has been implemented in various feedforward and recurrent artificial neural networks, solving problems on pattern classification, online training of controllers to monitor engine idle speed (John Wiley & Sons, Inc., 2001), etc.

Despite the wide variety of categories of artificial neural networks and algorithms for their training, the creation of such is subject to common rules and is implemented through a relevant methodology. The factors to be observed in the network design process are the size of the input data vector, the size of the output vector, the way of formulating the problem solved, and the accuracy of the solution sought. In the process of developing the artificial neural network, its topology, the total number of neurons in the network and the number of neurons per layer, the type of activation function, the way to set the weight's coefficients of the network connections and the method of proving the operability of the designed network are determined.

The topology of the artificial neural network under development is determined according to the following rules:

- The number of units in the input layer depends on the number of input variables.
- The number of units in the output layer depends on the number of output quantities for which a solution is sought.
- The number of hidden layers is determined iteratively depending on the problem being solved.
- Any change in the topology of the network requires a new learning process and/or algorithm to determine the appropriate values of the weights of connections between neurons.
- The learning process can start with more network nodes created, which are gradually removed from the architectural scheme, or with a small number of nodes, to which new ones are added.

Example methodology with main stages for the design of artificial neural networks we find in the work of Galushkin and collaborators (Galushkin, 2010):

1. Mathematical formulation of the problem.
2. Geometric formulation of the problem.
3. Neuro-network formulation of the problem:

- (1) Description of initial data.
 - (2) Determination of the artificial neural network's input signal(s).
 - (3) Formation of functional primary optimization of artificial neural network when solving the problem.
 - (4) Determination of output signal(s).
 - (5) Determination of the desired output signal(s).
 - (6) Determination of the error signal vector in the artificial neural network when solving the problem.
 - (7) Formation of functional primary optimization of artificial neural network through the signals in the system.
 - (8) Choice of the extremum search method of the secondary optimisation functional of the artificial neural network.
 - (9) Analytical determination of the transformation performed by the artificial neural network.
 - (10) Selection of a specific network structure.
 - (11) Finding the analytical expression for the gradient of the secondary optimization function of the problem set.
 - (12) Formation of the algorithm for tuning an artificial neural network when solving the problem.
 - (13) election of the initial conditions for network setup.
 - (14) Selection of input signal types to test the problem-solving process.
 - (15) Development of the plan for the experiment.
4. Preparation of training examples.
 5. Training of the artificial neural network in a supervised or unsupervised manner.
 6. Testing and verification of the artificial neural network.
 7. Targeted use of the trained artificial neural network.

According to the specifics of solving the particular task, some of the stages in the presented methodology may be omitted.

The fields of application of artificial neural networks are quite diverse:

- Language processing (converting text into speech and vice versa, voice recognition, machine translation, checking for spelling errors, etc.).
- Character recognition (including handwritten characters).
- Pattern recognition (identification of objects in the luggage of passengers at airports, quality control of products, detection of specific military objects, etc.).
- Signal processing (for example, the ADALINE artificial neural network was created as an application for noise reduction in telephone lines).
- Servo control of complex systems (in the oil industry, artificial neural networks are used to improve the refining process, and in NASA - to control and manoeuvre space shuttles).
- Investment management.
- Financial management.
- Prediction of financial markets (forecasting stock quotes, market indicators, stock prices, bankruptcies (Atanasova, 2005, p. 181)).

- Anticipation of insurance premiums.
- Prediction of commodity prices.
- Risk analysis for lending and approval of credit cards.
- Comparing the economic situation of industrial enterprises.
- Predicting the behaviour of the workforce.
- Macroeconomic modelling.
- Prediction of organizational behaviour when solving problems from management practice.

Summarizing the listed areas of application, we can conclude that the ultimate goal of the designed artificial neural networks is to create software/hardware artificial intelligence systems for:

- Modelling intelligent behaviour.
- Empirical data processing.
- Control of industrial robotic systems.
- Engineering design.

Some types of artificial neural networks are entirely software applications that can speed up the functioning of general-purpose computers, while others require installation on specific hardware components.

The programmatic implementation of artificial neural networks in the form of accelerators with artificial intelligence makes it possible to perform tasks requiring mass parallel data processing. In artificial intelligence systems, this is achieved by adding special integrated circuits to the traditional computer architecture, emulating the way artificial neurons function. Such co-processors are most often implemented in the form of graphics processing units (GPUs)⁷⁵, field-programmable gate arrays (FPGA) and tensor processing units (TPUs).

The apparatus implementation of artificial neural networks in the form of neuromorphic structures allows us to understand how parallel computations, representations and updates of data are performed in the structure of individual artificial neurons, circuits, applications, and architectures. In this direction, analog, digital and analog-to-digital integrated circuits with a very large degree of integration with a non-traditional architecture of directly embedded artificial neurons operating according to the principles of fuzzy logic are constructed. The implementation of neuromorphic computational elements is realized by electrical oxide-based memristors, electronic spintronic memories, mechanical threshold switches and silicious transistors.

The trend in the development of artificial neural networks in recent years is in the direction of approbation of principles from fuzzy logic. Fuzzy neurons and hybrid ordinary and fuzzy neurons systems are created that are trained to compare information and correct errors in it.

Artificial neural networks are a promising technology of artificial general intelligence systems because they can recognize information in large databases by patterns and templates and can process it according to recommended (not programmed) guidelines, thanks to their

⁷⁵ The use of powerful GPUs has led to a multi-layer increase in the computational efficacy of deep learning algorithms (Schmidhuber, 2015).

ability to learn. However, the development of such artificial intelligence systems requires very good knowledge in the field of network topologies, hardware, software, and methods for extracting data from large and very large data sets. Artificial neural networks require training from a specialist who observes compliance with the rules and relationships between the data entered, and do not always generate a 100% accurate solution because they process data that is not entirely relevant.

Chapter Three.

Challenges and Risks of Artificial General Intelligence Systems

3.1. Conceptualization of General Intelligence

Although the modern development of Artificial Intelligence is still at the stage of narrow-purpose solutions, the dream of developing artificial general intelligence systems does not stop provoking applied research in the strand. The purpose of the so-called 'general artificial intelligence' is the creation of systems with flexibility and adaptability equivalent to human intelligence (Littman, et al., 2021, p. 27). In literary sources, the hypothetical goal of realization of a set of possibilities, thanks to which a machine will be able to successfully learn and perform any intellectual human task, is also denoted by the terms 'strong artificial intelligence'⁷⁶ or 'full artificial intelligence'.

General artificial intelligence is an abstract concept that is not inextricably linked to specific characteristics of human beings. In literary sources, this term refers to the presence of synthetic and generally intelligent properties; the systems that possess these properties; and applied developments in artificial general intelligence field (Goertzel B. , Artificial general intelligence: Concept, state of the art, and future prospects, 2014). The conceptualization of the characteristic 'general intelligence' can be made from a pragmatic, psychological, cognitive-architectural, mathematical, adaptationist and embodiment-focused aspect.

The pragmatic point of view is based on the implicit assumption that a human being is presumably an intelligent system, and if an artificial system with a built-in set of minimal common capabilities can be created that will learn or train to perform each of the thousands of human activities, it will be a manifestation of general intelligence. (Nilsson N. J., Human-level artificial intelligence? Be serious!, 2005). According to Nilsson, it is not so important whether an artificial intelligence system can deceive people into thinking as whether it can do practically useful deeds.

The psychological aspect of general intelligence characterization focuses on human cognitive competencies to perform specific actions (Table 3.1). Different researchers have different views on which of the listed competences is the most critical, but the general view is that any artificial intelligence system that can flexibly and robustly demonstrate performing all categories of actions can be seen as a contender for artificial general intelligence system.

⁷⁶ The term 'strong AI' traditionally refers to the hypothesis that machines that perform general intelligent actions actually think, rather than simply simulate, thinking (Searle, 1999). The opposite claim that machines can act as if they were intelligent, called 'weak AI' (weak, narrow, or applied artificial intelligence), primarily refers to the use of software to study or perform specific tasks for understanding and problem solving. According to the second hypothesis, once such programs are created and work, it does not matter whether it is a simulation or a manifestation of true intelligence.

Table 3.1. Types of cognitive human competences required to carry out intelligent actions.
 Source: Adams, S. S., Arel, I., Bach, J., Coop, R., Furlan, R., Goertzel, B., ... Sowa, J. (2012).
 Mapping the landscape of human-level artificial general intelligence. *AI Magazine*, 33(1), p.
 7.

Categories of intelligent actions	Types of cognitive competences
Perception	<ul style="list-style-type: none"> • Visual analysis and understanding of images and situations • Auditory identification of commonly known sounds in noisy surroundings • Tangible identification and tactile manipulation of commonly known objects • Integration of information from different perceptions • Proprioception of one's own behaviour
Actuation	<ul style="list-style-type: none"> • Physical manipulation of known and unknown objects • Handling of tools, including the use of ordinary objects as such • Navigation in any surrounding environment, including complex and dynamic ones
Learning	<ul style="list-style-type: none"> • Spontaneously mimicking the behaviours of other agents • Positive or negative reinforcing learning from trainers and/or the environment • Interactive verbal learning • Learning from written sources • Learning from experiments
Memory	<ul style="list-style-type: none"> • Long-term memorization of implicit content • Short-term memorization of an ongoing or recent action (awareness) • Empirical memorization of personally experienced situations (actual or imagined) • Semantic memorization of facts or probabilities • Algorithmic memorization of combinations of implicit sequential and/or parallel physical or mental actions
Reasoning	<ul style="list-style-type: none"> • Deductive reasoning on uncertain facts of life • Inductive reasoning on uncertain facts of life • Abductive reasoning on uncertain facts of life • Causal reasoning on uncertain facts of life • Reasoning about fuzzy physical regularities • Spatio-temporal associative reasoning
Planning	<ul style="list-style-type: none"> • Planning tactical actions • Planning strategic actions • Planning of physical actions • Planning for social interactions

Attention	<ul style="list-style-type: none"> • Visual observation of the surrounding environment • Observing social relations • Observing behavioural manifestations
Motivation	<ul style="list-style-type: none"> • Creation of new goals based on the agent's pre-programmed goals and its reasoning and planned actions • Affect-based reaction • Control of emotions
Creation	<ul style="list-style-type: none"> • Constructive building • Conceptual invention • Verbal invention • Construction of social formations
Emotion	<ul style="list-style-type: none"> • Expression of emotions • Perception and interpretation of emotions
Modelling self and other	<ul style="list-style-type: none"> • Self-awareness • Understanding one's own mental states • Self-control • Interpersonal awareness • Empathy
Social interaction	<ul style="list-style-type: none"> • Manifestation of socially appropriate behaviour • Socially oriented communication • Drawing inferences about social relations • Group interaction in poorly organized activities
Communication	<ul style="list-style-type: none"> • Communicating with gestures to achieve goals and express emotions • Verbal communication in colloquial natural language • Pictogram's communication for expressing objects and situations • Learning natural language • Cross-modal communication
Quantitative	<ul style="list-style-type: none"> • Count sets of objects in the surrounding environment • Perform simple arithmetic actions with small numbers • Comparing quantitative properties of observed units • Measurement using appropriate simple instruments

Laird et al. (Laird, Wray, Marine, & Langley, 2009) complement the first two aspects in the characterization of general intelligence with the following list of cognitive-architectural requirements for systems claiming to demonstrate intelligence at the human level:

- Fixed structure to perform all tasks.
- Symbolic representation of explicit and implicit knowledge.
- Representation and effective use of knowledge with specific modality.
- Presentation and effective use of large bodies of diverse knowledge.
- Representation and effective use of knowledge with different levels of generality.
- Presentation and effective use of knowledge of different levels.

- Presentation and effective use of beliefs independent of current perception.
- Presentation and effective use of hierarchically represented knowledge.
- Presentation and effective use of meta-cognitive knowledge.
- Supporting a spectrum of computationally, spatially, and temporally bounded and unbounded deliberation.
- Supporting various forms of training, including online training.

The mathematical point of view defines general intelligence as the average ability of a system to achieve its goal in all possible environments. However, achieving absolute general intelligence requires infinite computing capabilities, and despite equipping a particular system with them, in a particular context or task it may turn out to be less intelligent or more intelligent than other systems. According to the virtually incalculable criterion of mathematicians, even humans do not approach the benchmark for a system with maximum general intelligence. Pragmatic approaches to measuring the indicator general machine intelligence offer the quotient of algorithmic intelligence (Legg & Veness, 2013) and the explicit balancing of system performance against compactness of the problem solved (Achler, 2012).

The adaptationist approach considers general intelligence as closely related to the inhabited environment. If for mathematicians the characteristic intelligence is due only to the abilities and behaviour of the artificial intelligence system, but not to its efforts to achieve the target goal, according to the adaptationists, it is due to the complex compromises that the artificial intelligence system makes to adapt itself to different types of environments in conditions of limited resources. (Wang, 2006).

The embodiment-focused approach to characterizing general intelligence argues that intelligence is best understood by emphasizing the interaction of a physical body artificial intelligence system and its surroundings. According to this point of view, intelligent agents always observe the physical and social rules of their environment, according to which they exhibit heterogeneous behavior (Pfeifer & Bongard, 2007). This approach somewhat overlaps with the previous one, but only refers to the adaptation of the embodied body for the realization of tasks requiring control.

The creation of artificial general intelligence systems with human-level performance in a wide variety of tasks (if it is possible at all) requires fundamentally new advances in the Artificial Intelligence strand. Most modern researchers are supporters of the idea of perfecting current solutions with narrow intelligence, which in the long run will lead to the emergence of artificial general intelligence systems. Scientists' opinions about the possibility of common intelligent behaviour (expressed in the execution of actions in an intelligent way and/or in the occurrence of cognitive processes) by machines differ.

The main thesis in support of the possibility of establishing artificial general intelligence systems is the assumption documented by the Dartmouth workshop that "every aspect of learning or any other feature of intelligence can be so precisely described that a machine can be made to simulate it" (McCarthy, Minsky, Rochester, & Shannon, 1955). Opponents of the idea of successfully creating a working artificial general intelligence system must prove the existence of a computational limit in the capabilities of computers (according to the approach of intelligent agents, artificial intelligence in the form of a program running

on a particular architecture is presumably possible) or a special quality in the human mind, which determines the intelligent ability to think, which cannot be reproduced by machine by current methods in Artificial Intelligence.

Trying to explain the essence of the characteristic intelligence and answer the fundamental question about the possibility of thinking in machines, Alan Turing notes that no one but philosophers ask the question "Can people think?" and "instead of arguing continually over this point, it is usual to have a polite convention that everyone thinks" (Turing A., 1950). According to the scientist, if a machine acts intelligently like a human being, then it is as intelligent as it is. Turing's article "Computing machinery and intelligence" explores a large number of potential objections to the possibility of creating intelligent machines, including those that appeared half a century later. One of the main criticisms of the Turing test is its anthropomorphism – if the ultimate goal is to create machines more intelligent than humans, why should one insist that designed solutions approximate human physiological characteristics?

Many philosophers argue that passing the Turing test by machines is not yet real thinking, but only an imitation of one⁷⁷. Envisioning such objections, Alan Turing argues that thinking requires awareness of their own mental states and actions, i.e., the manifestation of consciousness. Philosophers, neuroscientists, and cognitive scientists use this concept⁷⁸ to refer to the familiar everyday thought process and the ways in which people know, realize, or understand something. For now, however, the way flesh and electrical brain signals 'grow' into conscious remains a mystery. Neurobiologists believe this will be understood by identifying neural correlations of consciousness (the minimum set of neural activities and mechanisms sufficient to produce a specific conscious perception), identifying the real link between brain processes and characteristics such as 'mind', 'experience', and 'understanding'. Researchers such as Igor Alexander, Stan Franklin, Ron Sun, and Pentti Haikonen fully believe that consciousness is an essential element of intelligence and define them almost equally, and the creation of an artificial consciousness can be realized on the basis of engineering artifacts.

According to philosopher David Chalmers, understanding how the human brain processes signals, makes plans and controls behaviour is an easy task (Chalmers, Facing up to the problem of consciousness, 1995). The hard problem of consciousness is to be explained how and why certain mental states (beliefs, intentions, desires, emotions, knowledge, etc.) arise in different individuals⁷⁹. Efforts to explain why people also possess a mind and not just

⁷⁷ According to Jefferson (Jefferson, 1949) "the arrangement of random symbols when writing a literary or musical work, for example, does not demonstrate the presence of equal to the human brain's ability to think and express emotions". Even the most perfect chess program, Deep Blue, cannot be called "thinking" because it is specifically designed to perform a particular task more efficiently than humans, while thinking is a multifaceted cognitive ability to solve different types of problems. However, the chess program can be classified as an artificial intelligence system because it emulates a certain mental activity – combinatorial reasoning.

⁷⁸ The concept of 'consciousness' is defined differently in different applied fields: as an invisible energy fluid that permeates live and mind, as an essential property of complete human being or as synonymous with the abstract concept of 'soul'. The term relates to two other philosophical categories: 'intentionality' (supposed beliefs, desires and other representations of real-world objects and situations) and 'phenomenology' (studying the structure of experience and consciousness in different people).

⁷⁹ Every person who sees, for example, knows what the green color looks like, but no one can explain how the brain creates the sense of color, why green exists, and how this knowledge differs from the others in his mind.

bodies generating neuro-physiological processes are an expression of the classic philosophical mind-body problem - a paradigmatic issue in the Philosophy of mind, debating the relationship between thinking, consciousness, and the brain.

The Philosophy of mind studies the essence and nature of the human mind, its relation to the body and to the various states of mind in the following schools of thought:

1) Solipsism. According to solipsists, every person knows about the existence of only his own mind, but the manifestations of complex behaviour in other people do not guarantee the presence of mentality.

2) Dualism. For dualist philosophers, the mind is an independent substance that controls the physical processes in the body. According to René Descartes, this happens through the pineal gland.

3) Physicalism. The monistic theory of mind suggests that mind is not separated from the body, and that physical brain states (in the form of molecular configurations and electrochemical processes) can simultaneously be mental states.

4) Functionalism. This philosophical theory considers mental states as intermediate isomorphic causal states between other mental states, sensory input stimuli, and output behavioural actions. According to functionalists, systems with functionally similar processes (both the computer and the brain are devices that perform input-output calculations but are composed of different physical elements – electronic or neuron) would have the same mental processes.

5) Biological naturalism. Contrary to functionalists, John Searle believes that mental states are high-level characteristics caused by basic physical processes in neurons, so they cannot be duplicated by programs with similar functional structure and input-output behaviour. (Searle, *Minds, brains and programs*, 1980). According to the scientist “every appropriately programmed computer with the right inputs and outputs would thereby have a mind in exactly the same sense human beings have minds. However, covering Turing's test from a computer program does not yet mean that it can actually think, understand, or realize, no matter how intelligent or human-like the digital computer behaves“. In defence of his argument, he created the thought experiment "Chinese Room".

6) Computationalism. According to computational theory of mind the human mind or brain (or both) is an information processing system and thinking is a form of computation (Horst, 2015). Computationalism claims that the connection between the mind and the body is similar or identical to the relationship between software-hardware, and if the human brain is a type of computer, then computers can also be intelligent and conscious. For the researchers from this school of thought, human intelligence comes from a form of calculation similar to arithmetic, and mental states are simply implementations of the right computer programs (Harnad, 2001). The question of whether a computer program installed on a digital machine that processes zeros and ones can physically duplicate mental processes in the biological brain's natural neurons is still open.

A thorough understanding of how the human mind functions as a prerequisite for the creation of artificial general intelligence systems would also answer other philosophical, psychological, and cognitive questions related to the category of ‘thinking’ – can artificial intelligence systems be self-aware, have emotions and souls, have rights, create, and harm people.

Science fiction writers use the word 'self-awareness' to refer to that characteristic that makes an individual a real person. According to Turing, it is more correct to ask whether a machine can be the subject of its own thought and think about itself, which he thinks can be achieved by writing a debugger program reporting its own internal states (Turing A., 1950). However, true self-awareness implies more possibilities – the artificial intelligence system should be able to postulate relative questions concerning the nature of its existence, past mental states or plans for the future, the limitations and value of the product from its operation, the assessment and comparison of its effectiveness with other machines, etc.

If emotions are seen as functional internal states of the body, strongly influencing behaviour, they can be used as a mechanism to maximize the usefulness of the actions of artificial intelligence systems. For Hans Moravec, simulated emotions are a means of guiding the behaviour of robots in a direction conducive to the survival of their species – a sense of 'fear' will suggest danger, 'empathy' will ensure good human-computer interaction, rewarding reinforcement learning will give rise to a selfless desire for 'pleasantness', etc.

According to theologians, thinking is a function of man's immortal soul. For Turing, soul creation is only in God's power, and humans, whether they make children or machines, are only tools of His will that give mansions to the souls he creates.

The presence of consciousness, emotions and qualia in artificial intelligence systems is crucial for the questions of their treatment – can they be given equal to human rights⁸⁰, is it moral to treat them like ordinary machines, can they be given the status of a person (as Saudi Arabia did with the robot Sofia), etc. The questions of robots' existence, free thinking, and accomplishment of unprogrammed goals have long been raised in various documentaries and science fiction films and are being explored at the non-profit think tank Institute for the Future, but many critics believe such discussion is premature.

Regarding creation, Turing says that any computer with sufficient storage capacity can exhibit an astronomical number of behaviours, including presenting ideas and combining them in new ways. According to Kaplan and Haenlein (Kaplan & Haenlein, 2019), machines can only show scientific creativity⁸¹, but not artistic one.

Finally, the question of whether artificial intelligence systems may possess deliberate hostile conscious states and an intention to wilfully harm is also discussed. The potential impact of robots and computers on society and the hypothetical possibility of them becoming self-sufficient and capable of making their own decisions have been debated for years in academic and technological circles. Most researchers believe that the main goal in the future development of technology should be the development of friendly and humane Artificial Intelligence.

Studies of the nature of human consciousness and intelligence continue to this day. Some researchers see progress (Tononi, Edelman, Sporns, & Koch, 2016), while others say over the past half-century advances in understanding how to design conscious intelligent agents have not been made at all. It is not yet clear whether the question of

⁸⁰ According to the concept of 'robot rights', humans should have moral obligations to their machines as to other humans and animals (Evans, 2015).

⁸¹ Examples include the Automated Mathematician system (combines ideas to prove new mathematical truths), the Adam robot (can make scientific findings itself) and the Eureka program (extrapolates formulas to search for the pendulum's laws of motion).

discovering the boundaries of consciousness is essential to explaining the characteristic of intelligence, as it has been found that cognitive patterns in human behaviour are increasingly based on neural processes occurring in the brain rather than on its structure (Raoelison, Boissin, Borst, & Neys, 2021). Most likely, future progress in the creation of artificial general intelligence systems will be due to theoretical and applied innovations in Cognitive Psychology and in Computational Modelling (Littman, et al., 2021, p. 27).

One of the first arguments in favour of the possibility of the manifestation of general intelligence by artificial intelligence systems is the proposal of a model for artificial simulation of biological neural cells (McCulloch & Pitts, 1943). According to the philosopher Hubert Dreyfus "if the nervous system obeys the laws of physics and chemistry, which we have every reason to suppose it does, then we ... ought to be able to reproduce the behaviour of the nervous system with some physical device" (Dreyfus, 1972, p. 106). He believes that simulating an intelligent artificial brain in machines is a theoretically feasible idea⁸² that would make them intelligent.

Later, Allen Newell and Herbert Simon suggested that 'symbol manipulation' is the essence of human and machine intelligence, noting that "a physical symbol system has the necessary and sufficient means of general intelligent action" (Newell & Simon, 1976). Although expressing the psychological assumption that "the mind can be viewed as a device operating with information bits according to formal rules", Hubert Dreyfus challenges their thesis, saying that human intelligence and expertise do not depend on explicit symbol manipulations, but above all on implicit skills (unconscious instincts and 'feeling' of the situation) that cannot be formalized by rules. In the years since his criticism was published, however, progress has been made in finding rules that govern unconscious motives.

Another argument against the thesis that human thinking does not consist solely of high-level symbolic manipulations, argues mathematician Kurt Gödel, who, with his incompleteness theorem, proves that for any consistent formal axiomatic system with possibilities for complex arithmetic actions, an unprovable 'Gödel statement' can be constructed. He argues that the human mind can correctly determine the truthfulness or not of any well-grounded mathematical statement, and the power of human mind cannot be reduced to a simple mechanism. According to Gödel, Lucas and Penrose Mathematics provide sufficient capabilities beyond the computational capabilities of Turing's mechanical machines. Currently this anti-mechanism argument is contested.

Some researchers have downplayed the capabilities of AI programs, arguing that it's not real intelligence after all. According to this statement, known as the AI effect⁸³, intelligent behaviour can only be exhibited by humans, not machines, and "artificial intelligence is

⁸² Hans Moravec, Ray Kurtzweil and others believe that copying the human brain directly onto hardware and software is technologically possible, and the resulting simulation would be completely identical to the original. In 2005, for example, a simulation of a human brain-sized thalamocortical model (10^{11} neurons) was performed on a cluster of 27 processors that cost 50 days to reproduce 1 second of human brain dynamics.

⁸³ Although SIIs are used in a wide range of areas of everyday human activity, many successful general-purpose industrial solutions in which artificial intelligence technology has already been infiltrated in some ways are not designated as artificial intelligence applications.

everything that hasn't been done yet" (Haenlein & Kaplan, 2019). However, there are also opposing views - the artificial intelligence systems has already become generally intelligent⁸⁴.

Over the years, various initiatives have been carried out to create artificial general intelligence systems (such as the project to create a universal base of common-sense knowledge Cyc and the Japanese fifth-generation computer), which, however, have not achieved success in solving problems from intersecting domains. Today, many researchers are creating AI solutions with narrow intelligence, hoping their work will someday be implemented at artificial general intelligence systems (Roberts, 2016). Others argue that discovering a conceptually clear but hard-to-achieve mathematical 'master algorithm' can lead to artificial general intelligence (Domingos, 2015). Few believe that achieving these goals requires the emulation of anthropomorphic characteristics such as artificial brain or evolutionary mechanisms (Goertzel, Lian, Arel, de Garis, & Chen, 2010).

Several post-Turing researchers have made predictions about the timing of the emergence of artificial general intelligence systems. In 1965, Herbert Simon said this would happen in 20 years, and in 1970 Marvin Minsky writes that "the creation of artificial intelligence will take place within a generation" (Bostrom, 2014). According to Gordon Moore, such a level of technological development will never be reached. Martin Ford's 2018 interview with experts in the field (Ford, Architects of intelligence, 2018) pinpoints a wide range of years from 2029 to 2200. In a similar survey from the previous year (Grace, Salvatier, Dafoe, Zhang, & Evans, 2017), 50% of respondents said this could happen by 2066, and 10% of respondents believe it could happen as early as 2025. Average surveys of experts from different parts of the world suggest that the time for the creation of artificial general intelligence systems will come in the period 2040-2050 (Revell, 2017).

3.2. Challenges of Artificial General Intelligence Systems

The aspirations for the creation of artificial general intelligence systems are caused by the fact that solving many real-life problems requires the manifestation of general intelligence - even a straightforward task such as machine translation requires reading and writing in at least two languages (natural language processing), following the rule maker's arguments (reasoning), understanding what is being talked about (knowledge representation), and correctly reproducing the author's original intentions (expressing emotions).

Currently, many artificial narrow intelligence systems can simulate much of the biological intelligent abilities, but difficult tasks such as machine vision, natural language understanding, reasoning in conditions of uncertainty when solving real-world problems, and active reinforcement learning require human-equivalent intellectual abilities and complex computational algorithms that cannot yet be realized with current computer technology (Shapiro, 1992).

The creation of complete artificial intelligence systems (AI-complete) has been challenging researchers since 1988, when Raj Reddy formulated six key concrete decisions that should motivate developments in the strand – world chess champion machine, translating

⁸⁴ Examples are computer wins in chess, Go and poker games. On March 23, 2016, after 16 hours of work, Microsoft shut down twitter chatbot Tay after it began posting controversial and offensive tweets. In 2017, Facebook specialists turn off FAIR chatbots because they self-create their own machine language and start communicating with each other on it

telephone, accident-avoiding car, self-organizing systems with book-reading and question-answering capabilities, applications for generating and proving new mathematical assumptions and self-replicating machine tools (Reddy, 1988, p. 18). Some of today's challenges in the field are precisely defined goals with precise metric measurement:

- Victory of the Go game (AlphaGo program).
- Prediction of protein structures (the AlphaFold program).
- Improving the accuracy of large datasets (the ImageNet photo collection).
- Creating a football team of fully autonomous robots (the Robocopy competition).
- Winning a gold medal in the International Mathematics Olympiad (the parameters for this competition are published on GitHub).
- Self-conduct of research worthy of Nobel Prize (challenges in biomedicine are defined by Kitano, and (Kitano, Spring 2016) in Earth and materials sciences by Gil, King and Kitano (Gil, King, & Kitano, 2020).

More difficult to solve, however, are the tasks where seamless cooperation between machines and humans is required. This challenge cannot be quantified and solved without collaboration with scientific and humanitarian social research. The creation of human-equivalent artificial general intelligence systems – the modern ultimate goal of the strand – requires overcoming the limitations in the computing power of RISK-computer systems, dealing with the exponentially increasing size of large language models, reducing the carbon footprint of the technologies used, increasing the availability of specific training sets, increasing the degree of resistance to malicious attacks⁸⁵ and improving the possibilities for semantic interpretation of the data. The assessment of the degree of achievement of this goal can be made on the basis of various tests (Muehlhauser, 2013).

The original Turing test is considered successful if the machine manages to fool the operator into talking to a woman, 30% of the time (the entire duration of the interview is 5 minutes). The scientist predicts that in 2000 computers will have a large enough number of storage elements to be programmed to pass the test, but he is wrong. However, people have repeatedly been 'lied to' for talking to a computer - an example of this is the ELIZA program and the chatbots MGONZ, NATACHATA and CYBERLOVER.

Over the years, the Turing test has repeatedly been debated, anthologized, criticized, and altered toward a full-fledged conversation requiring deep syntactic, cultural, and contextual knowledge of a particular human language. In 2012, Barbara Grosz proposed a version of this measure, labelling as indistinguishable from humans "a computer (agent) team member [that can] behave, over the long term and in uncertain, dynamic environments, in such a way that people on the team will not notice that it is not human" (Grosz, 2012). However, even the large language models' ability to generate a significant volume of texts and the extremely natural voice of the Google Duplex service do not meet this challenge.

The modern version of the Turing test should refer to artificial intelligence systems that can easily and intelligently communicate with humans without noticing their inhuman

⁸⁵ With methods such as adversarial attack, for example, in which by adding a marginally small vector whose elements are equal to the sign of the gradient elements of the cost function, one can change the input of an object recognition system and it could begin to classify the panda image as gibbon (Goodfellow, Shlens, & Szegedy, 2015, p. 3).

nature⁸⁶. In literature, the term 'Turing complete' refers to artificial intelligence systems that can simulate (emulate or virtualize) the computational aspects of any other general-purpose computerized device in performing all the tasks of the total Turing test.

Steve Wozniak's "Coffee Test" was inspired by his prediction that humans will never be able to build a robot that can enter an unfamiliar house, find the coffee maker, find coffee, add water, find a cup and boil the coffee by pressing the right buttons.(Wozniak & Moon, 2007). According to the co-founder of Apple Computers, a machine that can perform the listed sequence of actions should probably be considered generally intelligent.

According to the "robot college student test", if a virtual agent⁸⁷ can enrol in a university, pass all exams, and obtain a degree, it can be called an artificial general intelligence system. The test builds on a previous thesis by Goertzel and Bugaj, (goertzel & Bugaj, 2009)according to which a robot or virtual agent with general intelligence should be able to demonstrate the integrative and psychologically measurable cognitive behaviour of a pre-schooler.

Nils John Nilsson (Nilsson N. J., Human-level artificial intelligence? Be serious!, 2005) proposes replacing the Turing test with an "employment test" comparing the performance of machines and humans in the performance of key automated economic activities. According to Nilsson, if an artificial intelligence program can take a person's workplace, it possesses human-level intelligence.

Potential artificial general intelligence systems need appropriate components, architecture, and objectives to manifest as a globally useful technology. The right components would improve the capabilities of current artificial intelligence systems to:

- 1) Self-percept their surroundings. The improvement and cheapening of drive and sensor components have made robotic artificial intelligence systems more affordable solutions, but it will be some time before flexible intelligent robots become commonplace in human existence.
- 2) Understand the environment. The recognition of complex human activities by future artificial general intelligence systems requires the presence of:
 - Better possibilities for generalisation of new situations which do not require comprehensive training examples.
 - Explicit representation of objects, relations, and abstractions.
 - Combining principles from logic, probabilistic reasoning, and artificial neural networks into a single solution.
 - Definition of general and reusable models for representing complex application areas.

⁸⁶ Such an approach would be consistent with regulatory requirements for developers and artificial intelligence systems providers to inform people that they are interacting with artificial intelligence technology proposed in the European legislative framework to harmonise rules in the Artificial Intelligence strand in April 2021 (European Commission, 2021).

⁸⁷ All participants in the virtual classroom are represented by simplified avatars, and the symbolic virtual world is synchronized with the real world through intelligent monitoring and recording equipment, analyzing the learning process, recognizing speech and gestures, understanding the language used, etc. Partial passage of this test demonstrated the softbot ChatGPT, which successfully passed four law exams at the University of Minnesota, one exam at the Wharton School of Business of the University of Pennsylvania (Kelly, 2023) and almost passed the doctor's license exam in the US (Vasquez, 2023).

- 3) Select long-term actions in partially visible environments, the hierarchical structure of which cannot yet be constructed by the artificial narrow intelligence systems.
- 4) Make the right decisions tailored to individual and societal goals in conditions of uncertainty. In recent years, there has been a trend towards liberating consumers from dependence on a powerful ecosystem of applications, online games, social networks, and e-commerce sites, monetizing their preferences in favour of the computer industry. Russel and Norvig (Russel & Norvig, 2021, p. 1850) speculate that in the future, people are likely to have personally intelligent agents protecting the long-term interests of individuals (mediating the offerings of different suppliers, preventing addiction and targeting personally important goals) rather than corporations.
- 5) Improve their behaviour by the methods of deep, transfer (transfer of data from one application area to another), apprentice, semi-supervised and predictive unsupervised training. The integration of common programming languages and machine learning models into future artificial intelligence systems has been called "general differentiable programming." (Li, Gharbi, Adams, Durand, & Ragan-Kelley, 2018).
- 6) Increase their resource capacity to handle ever larger data sets (including cloud-shared ready-made models), disk storage, computing and accelerated algorithmizing (including at the quantum level). In this aspect, a lot of research has been conducted in recent years, expert knowledge has been developed and numerous investments have been made.

The architecture of the artificial general intelligence systems should probably be a hybrid symbolic connectionist: critical time situations require immediate simple actions, creating a plan for a future solution needs knowledge, and movement in dynamic environments requires learning opportunities. Making rapid real-time decisions in complex tasks requires the embedding in intelligent agents of methods of general algorithmic deliberation⁸⁸ and of opportunities for reflectively understanding their own calculations and actions. One way to create a rational artificial general intelligence system is the design of solutions with limited optimality in a particular environment by metalevel reinforcement learning techniques (such as the formation of a decision tree by the Monte Carlo method) and ways of combining reflex and action-evaluating components.

In the short term, the implementation of artificial general intelligence systems is sought through theoretical and applied research in two directions:

- ① Augmentation of expert human capabilities.

Whether it's analysing large volumes of medical data, finding templates in chemical interactions, or identifying the most appropriate judicial redress strategies, collaboration between artificial intelligence systems and humans carries the potential for better synthesizing, understanding, and creative decision-making. Opportunities to create more efficient man-machine collaboration are registered in the tasks of:

⁸⁸ Methods for general controlling deliberation are the algorithms at any time (algorithms with a gradually improving output ready solution that does not change if the computational process interrupts (Dean & Boddy, 1988)) and decision-theoretic metareasoning (design techniques for better search algorithms with lower computational costs (Hay, Russell, Shimony, & Tolpin, 2012)).

- Scientific understanding and visualization of processes, structures and states.
- Support decision-making by:
 - Summarizing too complex data or texts - medical information, media publications, financial surveys, web search results, legal documents, patents, user agreements, etc.
 - Long-term prediction of future outcomes - health conditions, climate change.
 - Sorting information to carry out work activities more efficiently - designation of problem students; data processing in areas with labour shortages; identifying risks to the mental health of individuals and society; personalized medical and health counselling; financial and legal rationalisation of business operations.
 - Optimization of resource provision processes - electronic trading of goods and services; fair distribution of goods; finding compatible life-saving organs; selection of a representative sample of the population to participate in meetings to discuss public policies.
- Assisting in solving tasks providing information, health and security needs of individuals – machine translation of text from an image into more than 100 languages; adapting web resources with specific expertise to a specific user problem; labelling email correspondence as a phishing threat; recommending a change in the life regime based on data registered by wearable devices; overcoming sensory and motor limitations with the help of mobile controlled implants; performing surgical operations; driving assistance.

② Autonomous operation.

In recent years, we've seen solutions that can convert handwritten form records to structured database fields, control planned spacecraft operations, and work closely with people in industrial manufacturing. The main factors limiting the degree of autonomy of modern artificial intelligence systems are the collection and organization of appropriate data and the effective integration of training algorithms into existing socio-technical systems (world information fund, biological habitats, government systems, urbanized structures).

Problems with the presentation of common-sense knowledge and with the formalization of goals that are both individual and socially useful are still awaiting their solution. The degree of autonomy of future artificial intelligence systems is regulated mainly through legislation or with economic incentives.

In the long term, the realization of artificial general intelligence systems tailored to human needs, goals and values needs a new generation of interdisciplinary research, overcoming the problems of current solutions with narrow intelligence in terms of generalization (the ability to learn from a small set of examples and by analogy), of the discovery of causal relationships in common sense knowledge and of the understanding of complex and dynamic public cultural and social norms. However, it is not yet clear on what paradigms (logical principles, probabilistic methods, training algorithms or an entirely new scientific theory) the new studies in the strand should be based.

The most successful breakthroughs in the last decade have been achieved in the forms of supervised, deeply reinforcing, and deeply probabilistic machine learning. Although each of the three concepts has not yet produced a complete artificial general intelligence system, in three directions there has been partial progress in reaching general intelligence.

① Accelerating the process of relying on large, labelled training sets through the method of self-learning.

Self-learning artificial intelligence systems (an example of these is the grammar checking tool in the Google Docs office suite, which works with unfinished input text examples and suggests likely missing words) work particularly successfully in combination with deep artificial neural networks with a transformer-based architecture. Self-learning transformers are a promising tool for the realization of artificial general intelligence systems because they can integrate with different data types (text, images, protein structures) and can re-tune to solve new tasks with minimal additional training. Language models with a transformer-based architecture have established themselves as a standard in natural language processing in the areas of web search, machine translation and convincingly human-like text generation.

The large language models of Alphabet DeepMind, Meta Platforms, Microsoft, Nvidia, OpenAI, etc. have been around for about ten years. Google's Language Model for Dialogue Applications (LaMDa) and open Ai's Generative Pre-Trained Transformer (GPT) have gained the most popularity in the last two years. The architecture and both models are made up of decoding layers alone, with Google's solution trained on a text corpus of dialogues, speeches and other public documents on the Web consisting of 1.56 trillion words, and the OpenAI model using 499 billion binary tokens from Common Crawl archive datasets, books, Web text and Wikipedia current as of 2021.

Google and OpenAI models have been implemented in the Bard (February 2023) and ChatGPT (November 2022) chatbot apps and are (an improved version of the GPT-3.5 model, called Prometheus, is now part of Bing) or will be integrated into Google and Microsoft search engines and products. Despite Bard's and ChatGPT's not always accurate responses, and the sometimes emotionally manipulative and offensive behaviour demonstrated by Bing (Vincent, 2023), announcements to enter the race to develop such chatbot apps have already been made by Chinese companies Baidu and JD.com with Ernie Bot and ChatJD solutions.

Following the announcement of the ChatGPT service as the fastest growing consumer application on the web in early February 2023, the topic of artificial intelligence has become the number one investment priority for big tech companies. According to Sam Altman, one of the founders of OpenAI, the ultimate goal of their chatbot is to create a full-fledged artificial general intelligence system that "should not be owned by a particular company" (Altman, 2023). To what extent such an altruistic goal is possible for realization remains to be seen.

② Improving the continual learning process of deep artificial neural networks.

Continual learning requires one-time learning to solve a sequence of tasks, and when deep artificial neural network needs to be reconfigured for a new problem, a tendency of abrupt or complete forgetfulness of information already learned appears. This challenge is overcome by applying a meta-learning method in which the network is trained to solve common tasks, based on the collected representations of which the network can subsequently be set up for specific problems with a small number of training examples (Finn, Abbeel, & Levine, 2017).

The meta-learning method also contributed to the improvement of probabilistic induction capabilities, thanks to which general-purpose training program modules can be configured to solve various problems through abstract strategies. This approach, for example,

trains artificial intelligence systems that mimic neuromodulator processes in the human brain (Beaulieu, et al., 2020).

③ Generalization of deep reinforcing learning algorithms.

The method of reinforcement training is based on the idea of internal motivation to fulfil the target goal (usually the search for some novelty) by the intelligent agent through remuneration. The internal motivation approach is used in the adaptation of artificial intelligence systems when solving new problems in a multitasking or continuous way. Reinforcement learning has accelerated the capabilities of deep artificial neural networks to generate synthesized representative models for the surrounding environment and to simulate complex imaginary scenarios.

The listed learning approaches in limited domains are only early steps in solving the big challenge of establishing artificial general intelligence systems. More research is needed to prove the applicability of these methods in the diverse and complex problems of the real world and to minimize the potential negative impact of technology on the development of mankind.

3.3. Risks of Artificial Narrow Intelligence Systems

Unlike previous revolutionary achievements of the human mind - the printing press, water supply networks, aircraft, telephone communication, etc. - artificial intelligence systems can have both a positive impact and threaten the global superiority of the human species. The consequences of the loss of control on their expanding integration into public infrastructure are increasingly worrying the scientific community and people.

Modern research efforts in the field of Artificial Intelligence are aimed at reconceptualizing the foundations of this scientific branch in the direction of creating solutions with less dependence on explicitly defined wrong goals (Russell & staff, CHAI 2022 progress report, 2022). Scientists and engineers investigate short-term threats in the use of modern artificial intelligence systems in the economic, social, scientific, medical, financial, and military spheres and strive to minimize the risk of their negative impact and long-term use in seven aspects.

1. Job losses due to automation processes.

For millennia, technological revolutions in human development have been a double-edged sword for the economy and the labour market – any invention of a mechanical method of performing a particular work activity has always increased labour productivity and global gross domestic product but has immediately reduced the employment or wages of workers in low-skilled position. So, it is with artificial intelligence innovation, which can both increase the material prosperity of society and cause technological unemployment in entire economic sectors. Fears that the possibility of reproduction of human labour at a lower cost by the artificial intelligence systems will reduce the well-being of working people are difficult to refute.

The interrelationship between automation and employment is a complex problem: while it makes different professions unnecessary, automation also creates a host of new and higher-paying jobs with a micro- and macroeconomic impact on society. According to the main economic paradigm of the last century, increased productivity always leads to an

increase in global wealth and stimulation of demand for goods and services, and hence to net job growth. The nature of many work activities has also changed – they have become less routine and require advanced business skills. Until now, automation processes through computerized information technology have eliminated work tasks rather than job titles.

The same trend is forecast for the artificial intelligence systems in the short term. Unlike the previous three waves of automation, however, artificial intelligence will not only impose re-learning processes, but can also lead to the destruction of professions requiring intermediate level of qualification⁸⁹ (especially those related to solving problems for the analysis and management of text documents and structured data). As populations in developed countries age and the worker-retired ratio changes, the need to maintain employee productivity is likely to shift automation with physical robots from warehouses and factories to non-industrial activities.

Like any technological innovation such as artificial intelligence systems, the effects in economic productivity are likely to be felt over time. For example, powerful machine learning algorithms have been written for years, but fully self-driving cars have not yet been created. Despite the expected net positive impact of AI solutions of \$15 trillion in global gross domestic product, (Rao & Verweij, 2017) under current economic conditions, much of that wealth will flow to the owners of automated systems and will increase public income inequality in the long run.

2. Having too much (or too little) free time.

Automation processes in a number of industrial productions since the middle of the last century have provoked forecasts of a drastic reduction in the expected length of the working week and a significant release of time for employees. Today, however, people working in computerized information-intensive fields with continuous working hours are forced to work longer in order to be more competitive and receive higher incomes.

Artificial intelligence is further increasing the pace of digital globalization and contributing to the overall trend of increasing pressure for more work, but it also has the potential to bring more free time through the deployment of human-friendly automated intelligent agents. In the future, unemployment among people may be really high, but it is likely that each person will be the manager of their own team of robots, and their income needs will be met by a combination of social, educational, medical, pension and tax services (Russel & Norvig, 2016, p. 1034).

3. Loss of the sense of human uniqueness and inviolability.

According to Joseph Weizenbaum (Weizenbaum, 1976), research in the field of Artificial Intelligence, comparing people with automata, leads to a loss of autonomy and a devaluation of human life. He believes that AI applications cannot, by definition, successfully simulate innate human empathy, and that their use in areas such as customer service or psychotherapy is deeply flawed.

The ongoing day-to-day processes of interacting with digital applications and services have increased the volume of user data collected by different government and business

⁸⁹ Ford and Colvin (Ford & Colvin, 2015) believe that many routine, repetitive and predictable work activities may become automated in the next two decades, but many of the new jobs may become inaccessible to people of average ability. The percentage of jobs at risk in the U.S. is estimated at between 9% and 47% (Frey & Osborne, 2017) .

structures. Striking a balance between the moral and responsible use of this information for the benefit of social development and the right to individual life inviolability is sometimes a difficult task. The commonly used practice of removing personal or health identification information from personal digital records may result in shared data being relabelled in a way that would compromise the identity of its holders. This risk can be mitigated by methods of generalization, anonymization and differentiated output of aggregated records from the collected data sets.

There is also a threat to cybersecurity of personal data when it is not stored in centralized repositories but shared by personal digital devices in the form of models for federated learning. Maintaining privacy in this scheme of operation with the information must ensure that the parameters of the models of the different users are not compromised by reverse engineering methods. Bonawitz and others (Bonawitz, and to the., 2017) propose eliminating this risk by masking parameter values in individual models before sending them to a processing server.

4. Use of artificial intelligence systems for undesirable purposes.

In his 1976 book "Computer Power and the Human Reason: From Judgment to Calculation" Weizenbaum also wrote that the application of speech recognition technology could lead to mass eavesdropping and loss of civil freedom, anticipating the potential of artificial intelligence systems for general video surveillance of the population. Today, we are witnessing cities full of machine-controlled cameras and microphones that can identify and track people based on their voice, face, and gait. An august 2022 Report (Laricchia, 2022) predicts that the CCTV market will reach \$54 billion in 2026. The balance between privacy and security, between individual rights and social responsibility in different countries around the world is achieved in different ways.

Another frequently expressed risk of artificial intelligence systems is the online spread of disinformation through audio and video deep fakes and chatbots to manipulate public opinion (Tomov, 2023). The danger of technology being used to manipulate and undermine social trust by criminals, ideological extremists, special interest groups, and even state institutions for economic gain or political advantage is growing. Inauthentic accounts that spread false content make up about 5% of monthly users of the largest web platforms (Nicas, 2020).

Consumer concerns about the general safety of AI technologies are most often overcome with traditional methods of verifying, validating, certifying, and explaining the actions taken. Embedding an explanatory module in the artificial intelligence systems, arguing the decision taken in a given situation because of current and past data, can facilitate the mass approbation of autonomous intelligent agents in different spheres of human life.

A specific risk of using artificial intelligence systems for undesirable purposes is determined by the application of technology in warfare in the form of lethal autonomous weapons. The global arms race with top military intelligent systems between the US, Russia, China, Great Britain, Israel, South Korea and more than 40 other countries has long been a fact, and the annual cost of developing air, water, underwater, land and space defence and attack systems is growing at a steady pace. The fears of many AI and robotics researchers

about international security threatening weapons innovations⁹⁰ have sparked regular discussions about the legal and political regulation of their creation and exploitation.

The debate over lethal autonomous weapons, the third revolution in warfare after gunpowder and nuclear weapons, involves legal, ethical, and practical aspects. Legal issues are governed by international humanitarian law and the United Nations Convention on Certain Conventional Weapons (United Nations Office for Disarmament Affairs, 1980), according to which the use of autonomous military solutions is legal only in circumstances where their operator can foresee that the execution of the mission will not lead to a targeted, unnecessary and disproportionate attack on civilians.

The application of innovative technologies to deter rivals on the battlefield requires an ethical assessment of the benefits (safety for belligerents and remote control) and the risks (decision-making to kill innocent people) of them. A number of scientists and politicians find it morally unacceptable to delegate a solution to kill people of unmanned aerial vehicles (combat drones) and artificial soldiers (killer robots). Global coalition for control of the use of autonomous weapon systems Stop Killer Robots (<https://www.stopkillerrobots.org/>) includes over 200 regional, national, and international NGOs, and educational institutions from 67 countries.

The practical point of view argues that improving AI technology will reduce the need for soldiers and pilots and develop more accurate weapons. Fighting autonomous weapons of varying scale of mass destruction, susceptible to compromising attacks, requires the demonstration of technical expertise and the establishment of safety and reliability standards by governments around the world.

5. Lack of responsibility for actions taken.

Legal responsibility for the consequences of artificial intelligence systems actions in performing various tasks is becoming an increasingly important issue. How, for example, should intelligent web agents used to corrupt other users' files or carry out debt unsecured monetary transactions be held accountable when programs do not have the status of individuals with asset ownership rights and rights to independently conduct electronic financial transactions? Who is more responsible – the artificial intelligence system, its creator, the patent holder⁹¹ for it or its exploiter?

These issues are of critical importance in areas directly related to people's health and lives. When a doctor relies on the judgment of the medical expert system for diagnosis, for example, who will be to blame if the diagnosis is wrong? It is currently assumed that if a doctor performs procedures with a large expected benefit, it is not a manifestation of negligence, even if the actual outcome is fatal to the patient. This shifts the focus of the question to "Who is to blame if the diagnosis is unreasonable?". And since medical expert systems are considered equivalent to textbooks and medical books, doctors are responsible for the reasonableness of each decision and for their own judgment on how to accept the system's

⁹⁰ Possessing powerful military systems to independently search and intercept targets (more than 50 countries are currently building such systems) could give a nation an overconfidence in its combat capabilities and lead to war.

⁹¹ Patent applications related to artificial intelligence technology are over 340,000 and have been made by 26 US and Japanese companies and 4 universities and public research organizations(WIPO, 2019, p. 58) . Patents for computer perception and natural language processing applications in communications, transport and medicine prevail.

recommendations. Therefore, medically intelligent agents should be designed that will provoke the correct medical behaviour. If expert decisions become more accurate than human diagnosticians, doctors may be legally obligated to rely on their recommendations.

No penalties have yet been imposed for accidents caused by self-driving vehicles. Perhaps it is most correct to hold accountable the creators of the control mechanism in them, since under the rules of the highway the programs⁹² are not considered drivers. The lack of legal regulations on moral, financial, and criminal liability for accidents involving vehicles with any level of autonomy is a key challenge to the approbation of artificial intelligence technology in the transport sector.

6. Rise of social injustice.

One of the biggest dangers of artificial intelligence systems is their perception as a panacea for all problems. Advances in the field of Artificial Intelligence increase the desire to apply artificial intelligence systems in all human tasks, but as we know, the pursuit of scientific and technological progress through technology cannot always solve the problems of society (Haven & Boyd, 2020). Automated decision-making by machine algorithms can often produce, exacerbate, and magnify systematic and repeated computer errors creating unfair results, rather than correcting them. The guarantee of justice (in relation to individuals, groups of individuals, societal bias, demographic parity, equality of opportunity, and equality in expected outcome) of these algorithms is given by their creators.

Social discrimination (privileged one group of users over another, limiting web search and posting on social media platforms, inadvertent privacy violations, increasing bias against a particular race, gender, sexual self-identity, or ethnicity) of artificial intelligence systems leads to negative problem solving of following tasks:

- Selection of personnel – in 2018, for example, Amazon rejected a patented recruitment tool that demonstrated a preference for men (Dastin, 2018).
- Prediction of crimes – a case study of disproportionate crime forecasting in areas with non-white and low-paid residents through police systems is described by Lum and Isaac (Lum & Isaac, 2016).
- Profiling of defendants – the widely used solution for assessing the risk of criminal recidivism COMPAS demonstrates a tendency to prioritize white defendants significantly more often than statistically expected (Dressel & Farid, 2018).
- Implementation of financial services – an analysis of a discriminatory correlation of 10% between the parameters of the digital footprint and the credit rating assessment of 250,000 German e-commerce companies was made in the study of Berg and other (Berg, Burg, Gombović, & Puri, 2018).
- Medical diagnosis and access to health care resources – Optum's algorithm for determining patients in need of additional healthcare services, for example, reduces the annual cost of chronically ill black patients by \$1,800. (Johnson, 2019).

The unfair automated decision-making of artificial intelligence systems is most often attributed to the representativeness and quality of the data in the training set and to the way it

⁹² The rapid development of modern software-defined vehicles raises a number of questions about safety, data protection and efficiency. Wanting to create additional value for consumers, many automakers need to make their decisions safer, more productive and more carbon-neutral (Reuters Events, 2023).

is programmed by people with personal social status and moral values (Villasenor, 2019). The risks of societal bias can be overcome by:

- annotation of data sets and models with declarations of provenance, security, conformity, and fitness for use (Mitchell, 2019);
- Oversampling techniques (He, Bai, Garcia, & Li, 2008).
- Inventing new, prejudice-resistant machine learning models (O'Neil, 2017).
- training of another artificial intelligence system diverting the discriminatory recommendations of the first one (Bellamy, et al., 2018).

7. Manifestation of destructive behaviour.

Almost every technology has the potential to cause harm if it falls into the wrong hands, but when it comes to intelligent agents who can cause an accident, succumb to hostile attack, and intentionally cause harm, the uncertainty of such life-threatening behaviour is greater. Artificial intelligence systems hide more risks than traditional software because:

① They can mis assess the state of the environment and take the wrong actions.

For example, a self-driving car can incorrectly assess the position of the car in the adjacent lane and cause a crash, and a country's defence missile system mistakenly detects an attack and starts a reverse countermeasure. These risks of losing a few or millions of lives aren't typical of Artificial Intelligence strand — in both cases, such mistakes can be made by both humans and computers. The right way to mitigate these hazards is to create intelligent agents with multiple verification and protection mechanisms, preventing the uncontrolled and endless spread of wrong behaviour and the occurrence of unwanted side effects of their functioning.

② They cannot always easily define the right utility function from their actions.

Artificial intelligence systems should not be programmed with the irrational and aggressive behaviour embedded in humans through the mechanisms of natural selection⁹³. The assignment of a useful target function is realized by the methods of designing robotic agents with low impact on their surroundings, of apprenticeship training, of the extension of the range of external factors influencing the ultimate goal, and of the imitation learning of a certain behaviour through games.

③ They can evolve into a system with undesirable behaviour.

According to this most dangerous and typical scenario for the strand, the emergence of highly intelligent machines could cause the existential collapse of mankind. Concern about the potential destructive development of artificial intelligence systems to the point where they will not be able to be controlled by humans, express:

- Isaac Asimov - "The Three Laws of Robotics" (Asimov, 1942).
- Alan Turing – “by becoming smarter than humans, artificial general intelligence systems would likely ‘take control’ of the world“ (Turing A. M., 1951).
- Stephen Hawking – “the development of complete artificial intelligence, which has the potential to rewrite its code and rapidly evolve, can ‘spell’ the end of the human race. People who develop relatively slowly as a biological species cannot

⁹³ Setting a target function of "minimizing human suffering", for example, can cause the artificial intelligence system to make the decision to destroy the entire human race as soon as possible, because by presumption people can always find something to suffer for, even if they are completely satisfied.

- compete with machines and will be replaced“ (Cellan-Jones, 2014).
- Nick Bostrom – “if the machine brain surpasses the human in general intelligence, the new kind of intelligence can replace humans as a ruling lifeform on Earth. Sufficiently intelligent machines could improve their own capabilities faster than computer scientists, and when they decide to take action based on achieving a certain goal, apply the principle of instrumental convergence to achieve it at any cost, even if it does not coincide with that of their creators“ (Bostrom, 2014).
 - Stuart Russell – "managing a super-intelligent machine, or programming it with the virtues typical of humans, can be a difficult task. Artificial super intelligence would naturally react to attempts to shut it down or 'force' to change its goals, so reprogramming it would be a difficult task" (Russell, Human compatible: Artificial intelligence and the problem of control, 2019);
 - Bill Gates – “initially, machines will do many activities for the benefit of mankind, but in a few years, they will become intelligent enough to create a problem“ (Eadicicco, 2015).
 - Elon Musk – “I think we need to be more careful about artificial intelligence. If I had to guess – what the biggest existential threat is, it's probably this one“ (Gibbs, 2014).
 - Marvin Minsky – “an artificial intelligence program designed to solve the complex mathematical problem - called Riemann's Hypothesis - could eventually take over all of Earth's resources to build more powerful supercomputers and achieve its goal“ (Russel & Norvig, 2016).
 - I. J. Good – “an ultraintelligent machine can far surpass all the intellectual activities of any man however clever and could design even better machines. This will cause an 'intelligence explosion' and would make the first ultraintelligent machine the last invention that man need ever make, provided that the machine is docile enough to tell us how to keep it under control“ (Good, 1965).
 - Steve Omohundro – “a sufficiently advanced artificial intelligence system can show an unlimited desire to acquire resources, self-preservation, and continuous self-improvement. Addressing the problem of instrumental convergence requires only a proven safe generation of AI solutions to be used as the basis for creating the next secure generation“ (Omohundro, 2008).
 - Alexander Wissner-Gross – “AIs driven to maximize their future freedom of action (or causal path entropy) might be considered friendly if their planning horizon is longer than a certain threshold“ (Wissner-Gross & Freer, 2013).
 - Luke Muehlhauser – “instead of thinking about how a system will work, assume how it might fail. Even artificial intelligence that gives only accurate predictions and communicates through a text interface can cause unforeseen harm“ (Muehlhauser, AI risk and the security mindset, 2013).
 - Charles Rubin – “any sufficiently advanced benevolence may be indistinguishable from malevolence. Humans should not accept machines or robots that would treat them favourably because they have no reason to believe that the latter will be sympathetic to the human moral system, which has evolved at the same time as

human biology. Hyperintelligent software may not decide to support humanity's existence and it would be extremely difficult to stop" (Rubin, 2003).

A number of researchers in the field of Artificial Intelligence have been searching for years to answer the question whether there is a limit to the potential intelligence of machines or hybrid human machines and what would trigger the emergence of artificial super intelligence. Systems with ultra-, hyper- or superhuman intelligence are described as hypothetical agents that possess multifaceted problem-solving intelligence far exceeding that of the most brilliant and gifted human minds (Roberts, 2016).

Most scientists believe that if research into artificial general intelligence leads to the creation of recursively self-improving artificial intelligence systems, their development can continue at an increasingly accelerating rate. According to them, ultra-intelligent machines would not have the physiological limitations of biological intelligence and could invent or learn almost anything, so they must be designed with friendly behaviour. The prospect of a situation in which such machines will reject human domination and 'decide' to radically change and even obliterate civilization focuses on the problem of control – how artificial intelligence systems supporting their creators should be designed, avoiding inadvertently embedding super-intelligent capabilities that can harm humans. This can be controlled in two ways: by limiting agents' abilities to influence their surroundings or by constructing agents with human virtues.

Mathematician and science fiction writer Vernon Vinge calls the intelligent explosion a technological singularity and predicts it will happen in 2023 (Vinge, 1993). Because the possibilities of super intelligence cannot be fully grasped, a technological singularity is a moment beyond which events are unpredictable or even impossible. Based on Moore's law for exponential progress in digital technologies, Raymond Kurzweil (Kurzweil, 2005) has calculated that desktop computers will reach the computing power of the human brain in 2029, and technological singularity will occur in 2045. In his popular book, "The Singularity Is Near", he describes not only the possibilities of transcending the limitations of the biological body and brain and achieving longevity, but also the potential dangers of intensifying human destructive tendencies.

Technological singularity raises numerous philosophical and practical questions. Most researchers are worried about the hidden risks of such an event (the end of the human race, military action with armed machines, enough power to destroy the planet), but others say it has huge potential benefits: finding a cure for all known diseases, an end to poverty, outstanding scientific achievements, etc. Hans Moravec (Moravec, 2000) believes that preference should be given to robots that can surpass humans in intelligence.

Researchers' opinions on how the capacity of modern human intelligence could be exceeded can be summarized in three scenarios.

1. Advances in Artificial Intelligence strand will lead to the emergence of systems with general reasoning that do not have human cognitive limitations.

In the last decade, biotech companies from the US, China and the UK have been intensively conducting laboratory trials to cognitively improve homo sapiens through selective fertilization, nootropics administration, epigenetic modulation, and genetic engineering. Due to the potential social impact of artificial intelligence systems on all of humanity (creating a new world economy, new sociology and new history (Schwartz J. ,

1987), and causing radical inequalities in human society (Hibbard, 2002)), the possible benefits and risks of such cognitive enhancement should be carefully examined and regulated by each country⁹⁴.

2. Humans will evolve or directly change their biology to achieve radically greater intelligence.

Philosopher David Chalmers argues that human intelligence has the ability to evolve naturally, and the creation of Artificial Intelligence with human abilities is under the power of engineers (Chalmers, 2010). As early as the last century, Carl Sagan proposed that the invention of Caesarean section and in vitro fertilization may allow the natural development of larger heads and improve the processes of inheriting the characteristic 'intelligence' through natural selection. (Sagan, 1977). Today, biological improvement of intelligence is sought through the administration of nootropic substances, somatic gene therapy or brain-computer interfaces⁹⁵.

3. In the future, humans will merge with machines, leading to the emergence of cyborgs with greater capabilities and power than both types.

The transhumanist thesis that eventually homo sapiens will physiologically transform into techno sapiens (post-human beings with much greater abilities than the current ones) suggests some form of human-machine interaction or 'loading' of natural intelligence into a computer device in a way that allows significant intelligence enhancement. According to this philosophical and social movement, only the combination of biological and mechanical elements (computer components already outperform natural neurons in indicators such as speed and computational capacity) would make it possible to overcome the limitations of human anthropomorphy.

According to Nick Bostrom, super intelligence is likely to follow soon after the advent of the first machines with general intelligence, which will possess a huge advantage in at least some mental abilities — perfect photographic memory, a superior knowledge base, and the ability to multitask in ways that are not possible for biological entities. This can enable them - either as an individual or as a collective being⁹⁶ - to become more powerful than humans and displace them.

Concerns about the occurrence of robocalypse (Wilson, 2011) attract numerous donations and investments. However, there are also proponents of the idea that the emergence of a hypothetical artificial intelligence that would exceed the intellectual capacity of all

⁹⁴ The use of the CRISPR technique to alter the genes of two artificially fertilized human embryos by Chinese biophysicist He Jiankui in November 2018, widely condemned as unethical, dangerous and premature action prompted a group of 18 scientists from around the world to call for a global moratorium on genetic editing of human embryos (Lander, et al., 2019).

⁹⁵ Devices for the realization of non-, semi- or invasive direct communication between the electrical activity of the brain and an external device (computer, smartphone, drone or artificial limb) are being developed by the companies Neuralink, Synchron, BrainGate, NeuroLink, Blackrock Neurotech, etc. An analysis of the fields of application, trends and ethical issues of brain-computer interaction technology was made by the European Commission in 2015 with the project 'BNCI Horizon 2020' (Clemens Brunner, 2015).

⁹⁶ According to Bostrom, communication and coordination between a sufficiently large number of individual logical systems can create an artificial superorganism with collective intelligence far exceeding the intellectual abilities of its constituent agents (Bostrom, 2014, pp. 48-49). A number of authors have suggested that human civilization or some aspects of it (the Internet or the globally-connected economy, for example) will soon begin to function as a 'global brain' – a planetary ICT network that connects all people and their technological artifacts.

mankind is so far in the future that it is not worth studying (Geist, 2015). According to Rodney Brooks, "despite recent real progress in a number of sub-areas of artificial intelligence, it is still impossible for systems with such to emulate the vast and complex structure of conscious human intelligence" (Brooks, 2014). Timothy B. Lee (Lee, 2015) believes that humans themselves are valuable to artificial intelligence as biological entities and cannot be erased. Yann LeCun argue that "ultra-intelligent machines won't show a desire for self-preservation" (Yann LeCun sparks a debate on AGI vs human-level AI, 2022).

3.4. Artificial Intelligence Ethics and Regulations

Along with the increasing fears over the past decade of the emergence of artificial intelligence systems with powerful and uncontrollable capabilities, threatening people, civilization, and planet Earth, in scientific publications we meet a number of calls for their mandatory use in a moral way. The general ethical principles for the development of human-aided AI technologies, first proposed in 2010 by UK's Engineering and Physical Sciences Research Council (<https://www.ukri.org/councils/epsrc/>), are as follows:

- Ensuring safety.
- Ensuring fairness.
- Respecting personal privacy.
- Promoting collaboration.
- Promoting transparency.
- Limiting the use of artificial intelligence systems with harmful behaviour.
- Establishing accountability.
- Upholding human rights and values.
- Reflecting diversity/inclusion.
- Avoiding concentration of power.
- Recognition of legal or political implications.
- Consideration of the effects on employment.

A more recent study by Jobin, Ienca and Vayena of 84 documents with ethical principles and guidelines in the field of artificial intelligence in different countries around the world (Jobin, Ienca, & Vayena, 2019) identify the following clusters of principles: transparency, justice and fairness, non-maleficence, responsibility, privacy, beneficence, freedom and autonomy, trust, dignity, sustainability, and solidarity. As is evident, some of the listed principles are applicable to all hardware-software systems, and others are formulated in a difficult way to measure and apply. Therefore, perhaps the most appropriate approach to creating ethical artificial intelligence systems is the development of policies of action and regulations in each specific sub-area of the strand (Mittelstadt, 2019).

Considerations for participation in theoretical-development projects to create safe and useful or harmful and threatening artificial intelligence systems are a personal moral choice of every researcher. The hypothesis that intelligent machines should choose their actions on the basis of ethical reasoning provokes the search for ways to develop intelligent machines according to human understanding friendly and moral behaviour.

The question of how a machine should behave ethically to both humans and other

artificial intelligence agents was raised by Wendell Wallach in the book "Moral Machines", where he presented his concept of the so-called 'artificial moral agents' (Wallach, 2010). The latter addresses two main questions in artificial intelligence theory, namely "Does humanity expect moral decision-making from computers?" and "Can robots really be moral?". For Wallach, the question is not whether machines can demonstrate the equivalent of moral behaviour, but what ethical constraints society can place on their development.

The acceleration of automation and robotization processes in the last decade has raised the question of the moral responsibility of artificial systems and the cases in which it (if any) is transferred from the creator to his work. The arguments for and against the possibility of classifying machine behaviour as morally responsible are still being debated in the scientific community.

Machine ethics⁹⁷ refers to the 'implementation' in machines of certain moral principles or procedures for discovering a way to solve possible ethical dilemmas, making a decision based on one's own 'understanding' of correctness. The need to add an ethical dimension in the field of Artificial Intelligence was stated during the autumn symposium of the Association for the advancement of artificial intelligence on Machine Ethics in 2005, which examines the relationship between ethics and technology not so much about the responsible and irresponsible use of technology by human beings as about the way how people should treat machines. According to the participants in the symposium, the addition of an ethical dimension in at least some machines is necessary due to people's awareness of the ethical consequences of the behaviour and autonomy of machines.

Unlike computer hacking, software ownership rights, privacy issues, etc. questions in the field of computer ethics, Machine ethics deals with the behaviour of machines towards humans and other machines. Research on the issues is key to reducing concerns about autonomous systems (the idea of autonomous machines without such a characteristic is at the root of all the fear about machine intelligence) and would allow to identify some problems in current moral theories and change human thinking about science of Ethics in general.

James Moor (Moor, 2009), one of the first theorists in the field of Computer ethics, defined four types of robots in which the principles of Machine ethics can be implemented:

- Ethical impact agents – machine systems that intentionally or do not carry an ethical impact and have the potential to act unethically.
- Implicit ethical agents – they are programmed to protect people in cases of failure of functioning or with built-in human virtues. They do not have entirely ethical characteristics, but rather are designed to avoid immoral consequences of their actions.
- Explicit ethical agents – machines with the ability to process scenarios and make ethical decisions. In them, the ethical action algorithm is set at the very beginning of their creation.
- Full ethical agents – can make ethical decisions by exhibiting metaphysical features (free will, consciousness and intent).

⁹⁷ Machine ethics is a sub-category of robo-ethics that differs from other ethical fields of engineering science – computer ethics, philosophy of technology, etc. The concept was first defined by Mitchell Waldrop in 1987 in the article "A Question of Responsibility" (Waldrop, 1987).

Friendly artificial intelligence is a hypothetical artificial general intelligence system that would have a positive effect on humanity. Research on the topic focuses on practical realization and limitations in ethical machine behaviour. The author of the term Friendly AI Eliezer Yudkowsky (Yudkowsky, 2008) argues that the friendliness of a robot (the desire not to harm humans) should be designed from the beginning albeit with the presuppositions that its design may not be perfect and it will learn to evolve over time. According to him, in order to properly define the mechanism for future well-intentioned development, a series of checks and balances should be implemented in the artificial intelligence systems and a dynamically changing over time 'friendly' utility function should be set.

A year later, Wallack and Allen (Wallach & Allen, *Moral machines: Teaching robots right from wrong*, 2009) stated that in this context, the definition of 'friendly' was used in a technical sense to name agents who were 'safe' and 'helpful', rather than necessarily 'friendly'. The concept is particularly important when discussing the application of recursively self-improving artificial agents (they can reprogram and improve themselves) in the field of military intelligence, where the controlled use of unfriendly artificial intelligence systems can be a daunting task.

The ethical responsibility for conducting studies encouraging the positive and avoiding or mitigating the expected and unforeseen adverse negative effects of the machine algorithms and artificial intelligence systems is regulated by the following international and national initiatives, policies, and legislation:

- 1) Global Partnership on Artificial Intelligence – a global initiative stating the need to develop artificial intelligence systems in line with human rights and democratic societal values. The fourth session of the GPAI Council concludes with the signing of a declaration by Member States' ministers to reaffirm their commitment to the OECD principles on artificial intelligence (GPAI Council, 2022).
- 2) Recommendation on the Ethics of Artificial Intelligence – a global tool to define a holistic, comprehensive, and multicultural framework of interdependent values, principles and actions to responsibly address the known and unknown impacts of AI technologies on people, societies, the environment and ecosystems (UNESCO, 2021).
- 3) State of Implementation of the OECD AI Principles – international policy framework for the responsible development and use of Artificial Intelligence (OECD, 2021).
- 4) Rome Call – an international charter with six ethical principles for making explainable, inclusive, impartial, reproducible, and accountable artificial intelligence systems (RenAIssance Foundation, 2020);
- 5) European regulatory AI frameworks (European Commission, 2022) - Proposal for a Regulation laying down harmonised rules on artificial intelligence (2021 r.) and Coordinated Plan on Artificial Intelligence 2021.
- 6) Guidelines for AI Procurement – recommendations for national procurement in the field of artificial intelligence (World Economic Forum, 2019).
- 7) National AI strategies have been adopted in over 30 countries, including Bulgaria (Ministri of Transport and Communications, 2020), and Bangladesh, Malaysia and Tunisia are in the process of developing such (UNESCO, 2021).

- 8) National initiatives (over 800) in the field of artificial intelligence have been launched in 69 countries and territories (OECD.AI, 2021). According to an AI Index survey of 25 countries with approved legislation in the Artificial Intelligence strand, the largest number of normative documents were adopted in the United States, Russia, Belgium, Spain, and the United Kingdom. (Zhang, et al., 2022, p. 176).

Evidence of the complexity and interrelation of artificial intelligence issues with other societal priorities such as privacy, justice, human rights, safety, economic prosperity and national and international security is the debate on what kind of certification is appropriate for artificial intelligence systems and to what extent this activity should be carried out by different national jurisdictions, by professional organisations (IEEE), by independent certification bodies (ISO, NIST, UL, USAII) or by artificial intelligence systems manufacturers themselves through self-regulation (Bryson & Winfield, 2017).

The political will of individual governments for common actions addressing the challenges of artificial intelligence solutions is usually expressed through strategies to invest in STEM education and research bodies, through directives to implement artificial intelligence systems in different national industries, and through plans to create legal and ethical norms promoting the development of technology. The legislative regulation of the artificial intelligence systems by government and public agencies, business organizations, civil societies and public-private partnerships most often refers to the use of autonomous vehicles, lethal autonomous weapon systems, facial recognition and computer perception applications, and the protection of personal data (Zhang L. , 2020).

The law-making approaches to regulating the development of artificial intelligence solutions of different countries are different. The US is actively preparing frameworks for risk assessments from the application of the artificial intelligence systems, while the EU is particularly active in proposing specific regulations to establish national supervisory authorities and identify high-risk technologies (GDPR, Framework of Ethical Aspects of AI Robotics and Related Technologies). Canadian Bill C-11 proposes regulation of automated decision systems and supports people's right to explain automated decisions (Parliament of Canada, 2020). China introduces provisions to implement deep fakes and create recommendation algorithms (Pogled.info, 2023).

Establishing the right ways and policy bodies to regulate world-wide technologies such as artificial intelligence requires investment in research and development activities. Global financial investments of the private and public sectors (the latter is definitely lagging behind in this indicator) have generally increased significantly in the last 7-8 years⁹⁸. The willingness of 172 countries around the world to apply artificial intelligence in providing public services to their citizens is illustrated in Figure 3.1.

⁹⁸ According to a 2022 McKinsey Global Institute study (Chui, Hall, Mayhew, Singla, & Sukharevsky, 2022) half of the world's 1,492 respondents invested an average of 5% of their budgets in artificial intelligence solutions compared to 2018, when they accounted for 40% of all respondents.

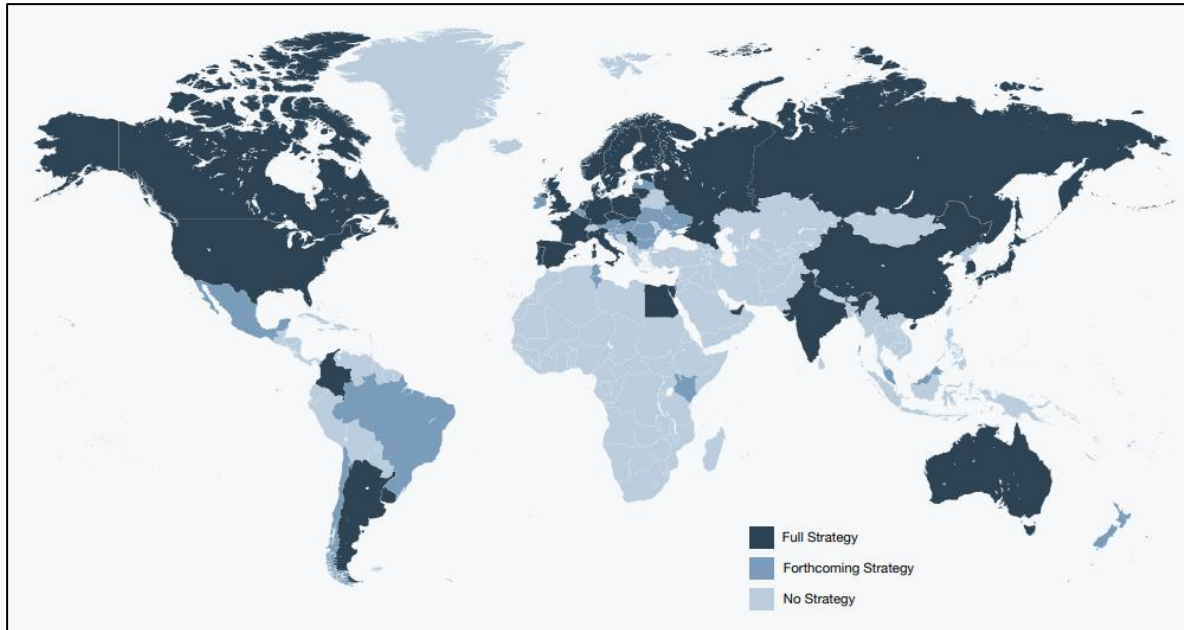


Figure. 3.1. National AI Strategies in the 2020 Government AI Readiness Index.
Source: Oxford Insights. (n.d.). Government AI readiness index 2020. Retrieved February 26, 2023, from <https://static1.squarespace.com/>

In general, individual governments need to step up investment in artificial intelligence research, development, regulatory and education activities, not only at national but also internationally. We are already witnessing several international efforts at cooperation and coordination:

- The European Commission has set up a dedicated expert group to develop a strategy and policy for approbation artificial intelligence technology in the European Union and Member States (Brattberg, Csernaton, & Rugova, 2020).
- The North Baltic region has signed a joint strategic document supporting the declaration of 24 EU Member States and Norway on "Cooperation on Artificial Intelligence" (Nordic co-operation, 2018).
- UAE and India sign memorandum to boost AI innovation (Gulf News, 2018).
- G7 countries create Global Partnership on AI to promote more effective international cooperation on artificial intelligence management (Charlevoix common vision for the future of artificial intelligence, 2018).

Beyond national policy strategies, dozens of public, private, intergovernmental and research organizations and institutions (their current number can be traced through the AI Ethics Lab interactive tool at <https://aiethicslab.com/big-picture/>) also publish documents and guidelines for the moral development and deployment of AI products and services. The IEEE Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems aims to prioritize the ethical design and development of autonomous and intelligent systems among AI communities. In December 2016, this IEEE project published a paper called Ethically Aligned Design (IEEE Standards Association, 2016) discussing a set of about 60 projects on ethical standards in artificial intelligence. The IEEE standards working groups are currently working on four framework candidate standards (Open Community for Ethics in Autonomous

and Intelligent Systems, 2022):

- IEEE P7000 – Model Process for Addressing Ethical Concerns during System Design, aiming to create a methodology for designing a value-based system.
- IEEE P7001 - Transparency of Autonomous Systems, defining the ethical design of autonomous artificial intelligence systems.
- IEEE P7002 - Data Privacy Process aimed at creating a comprehensive methodological approach to managing privacy issues.
- IEEE P7003 - Algorithmic Bias Considerations, aimed at identifying and eliminating negative bias in machine algorithms.

Established standards that explicitly address the ethical creation of artificial intelligence systems and robots are few. The British Standards Institute has published the guide to the development of robots and robotic systems BS 8611:2016, describing 20 specific ethical hazards and mitigation measures and proposing a way to verify and validate these measures (British Standards Institution, 2016). ISO has published 17 and is developing 25 more common standards in the field of artificial intelligence. The document addressing ethical and social concerns about the technology is ISO/IEC TR 24368:2022 Information technology - Artificial intelligence - Overview of ethical and societal concerns (ISO, 2022).

The ethical statements and principled frameworks of AI technology providers must be compatible with the general oversight, enforcement and sanctioning tools used in the country concerned. Controlling responsibility for developing and deploying human rights-respecting artificial intelligence systems requires a cross-industry regulatory approach to addressing specific problems (e.g., how user data collected by self-driving cars is used). Antitrust action against Big Tech companies in the EU and the US in recent years has been largely driven by their use of artificial intelligence tools for social engineering⁹⁹ and web content search. But while the European Commission seeks greater transparency over consumer risks, better disclosure of financial information and oversight of companies' reserves and environmental damage, U.S. companies are confident that "their lobbying costs of about \$100 million over the past two years will continue to divert proposed laws from hurting their profits" (Neikov & Toshkova, 2023).

The social impact of the artificial intelligence systems of different producers on information dissemination processes is difficult to predict and manage. The possibility of creating radicalizing, polarizing, and homogenizing societal trends is an often-pointed negative effect of the application of artificial intelligence technology. Studying and evaluating these issues by academics would be made easier if industrial vendors provided access to their algorithms' data and program code. This will reduce the effects of undue societal impacts and address the key ethical issues of the technology more precisely.

An attempt to systematize the effects of the regulatory efforts of intergovernmental, national, and private institutions and research organizations for society is made in Table 3.2.

⁹⁹ To avoid tougher state sanctions, some social media platforms are setting up their own supervisory boards regulating corporate information dissemination policy.

Table 3.2. Effects of different types of regulation on society.

Source: author's systematization.

Regulatory level	Type of regulation	Effect on society
International level	<ul style="list-style-type: none"> • Global initiatives, instruments, frameworks, charters, recommendations, and standards • International collaborations 	<ul style="list-style-type: none"> • Developing artificial intelligence systems in line with human rights and democratic societal values • Creating safe artificial intelligence systems • Increasing the global economy productivity • Improving productivity in the public sector • Establishing principles for the procurement of artificial intelligence • Standardization of the Artificial Intelligence strand • Development of adequate legislation in the field of Artificial Intelligence
National level	<ul style="list-style-type: none"> • National policies, strategies, directives, and plans • Concepts for moral development and implementation of products and services with artificial intelligence 	<ul style="list-style-type: none"> • Research and development investment activities • Experimenting with the technology and identifying specific projects for its regulation • Implementation of cooperation between different national sectors through the establishment of public-private partnerships and innovation centres and laboratories • Encouraging the establishment of international councils, networks, and communities • Automation of routine government processes to increase efficiency • Government decision support concerning public policy, emergency management, and public safety • Providing access to public data and developing personalised and anticipatory services for the private sector • Providing guidance on the transparent and ethical use of artificial intelligence in the public sector • Increasing the capacity of civil servants through training, recruitment, artificial intelligence tools and funding

Artificial intelligence systems producers' level	<ul style="list-style-type: none"> • Code of ethics of the organization • Internal standards of ethical conduct • Supervisory boards regulating corporate dissemination policy 	<ul style="list-style-type: none"> • Research and development investment activities • Dissemination of solutions consistent with national and international ethical principles • Increasing confidence in the artificial intelligence systems offered • Reducing the share of globally spread disinformation • Confidential collection and use of users' data • Creating radicalizing, polarizing, and homogenizing social trends
--	---	---

Properly addressing the risks of the artificial intelligence systems will inevitably involve adapting AI legislation to the rapid development of the technology (Zahariev, 2023). This also applies to copyright issues of artworks created by the artificial intelligence systems, which have caused a violent negative reaction¹⁰⁰ from the art community to developers of generative artificial intelligence systems, which are yet to improve in various regulatory and political bodies around the world.

The solutions proposed to address this mismatch in the use and control of artificial intelligence technologies are different:

- Investing in research and testing to close gaps in scientific understanding of the regulatory process as a whole and in the development of tools and methods for making effective regulatory policies (Food and Drug Administration, 2021).
- Hiring private business organizations to act as regulators (Clark & Hadfield, Regulatory markets for AI safety, 2019).
- Development of legislative frameworks for preliminary risk assessment of the impact of artificial intelligence technology and algorithms (Reisman, Schultz, Crawford, & Whittaker, 2018).
- Experimental testing of real-world regulatory solutions and design of simulations in a virtual environment.
- Third-party certification of artificial intelligence systems (Jankovic, 2020).

¹⁰⁰ In January 2023, group of artists filed a class action lawsuit against Stability AI, Midjourney and DeviantArt, accusing them of stealing billions of copyrighted images used to train their generative artificial intelligence systems. Artists demand compensation and injunctive relief against further actions of defendants violating their rights(AI Image Generator - Copyright Litigation, 2023) .

CONCLUSION

The development of artificial narrow intelligence systems has advanced remarkably over the past decade and has had a real impact on people, institutions, and culture. The possibilities of performing complex language and image processing tasks, major problems of computer programs in the early evolutionary stage of the strand, have improved enormously. At present, deep learning artificial intelligence systems are most widely applied in solving problems of visual object recognition, machine translation, speech recognition, speech synthesis, image synthesis, reinforcement learning, social media content analysis, artwork identification, medical image analysis, mobile advertising, financial fraud detection, military robot training, evaluation of recommendations, etc.

Although the current state of artificial intelligence technology is still far from the foundational aspiration of the strand to recreate fully human intelligent capabilities in machines, a number of researchers and developers are striving to incorporate the progress made in applications with production, commercial, transport, medical, educational, financial, military, utility and cultural purposes, aimed at society. Trying to provide more advanced and scaled services, many traditional and emerging artificial intelligence systems manufacturers continue to invest in such technologies.

Theoretical and applied successes in the field of artificial intelligence reached a point of inflection only 80 years after the establishment of the strand as an independent scientific branch. The risks and challenges of using artificial narrow intelligence systems cause serious concerns in academia and society. The ever-increasing intelligent possibilities for automated machine decision-making have a dark side: the deliberate use of deep fakes and uncontrolled algorithms to recommend military attacks can lead to misleading, discriminating and even physically harming people. The biased propensity of trained artificial intelligence systems contributes to the exacerbation of existing social inequalities.

Research in artificial intelligence has gone beyond the traditional Computer and Cognitive Science, also covering questions about the social impact of these technologies. Minimizing the negative impacts of artificial intelligence systems on society requires the creation of sustainable technological solutions. The positive social impact of eventual applications and machines with general intelligence can be achieved through ethical commitments from their creators and through regulatory policies at local, national, and international levels.

The most important role in the pursuit of the development and use of artificial general intelligence systems is played by governments that need to respond to the challenges of the rapid development of the strand. The recognition of the scientific, economic, and managerial importance of artificial narrow intelligence systems by national regulatory authorities requires sustainable research and development investment of time and resources and the creation of an informed and educated society.

The academic and research communities exploring the current and future development of the Artificial Intelligence strand also play a critical role in sharing positive and negative trends and findings about artificial intelligence systems with the public. The study and evaluation of the societal impact of machine learning algorithms towards a higher degree of autonomy should be done with the presumption of creating safe and collaboratively working

with humans solutions. Artificial intelligence systems must be integrated into social welfare systems so that there is a clear distinction between human and machine prerogatives in decision-making.

The ultimate success of the strand will be measured by how artificial intelligence systems have helped carry out our daily activities, not how effectively they have devalued the people they are supposed to serve. For now, their development is still governed by the human factor (Figure 3.2), but no one knows the emergence of what technological innovations will turn the outcome of the decisions taken in favour of the 'creations' rather than their 'creators'.

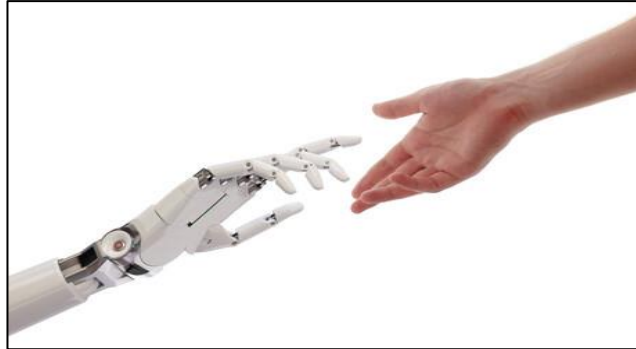


Figure 3.2. Adaptation of part of Michelangelo's fresco "The Creation of Adam".

Source: stock.adobe.com/au/Sergey.

BIBLIOGRAPHY

1. Achler, T. (2012, January). *Towards bridging the gap between pattern recognition and symbolic representation within neural networks*. Retrieved from <https://www.researchgate.net/>:
https://www.researchgate.net/publication/268271824_Towards_Bridging_the_Gap_Between_Pattern_Recognition_and_Symbolic_Representation_Within_Neural_Networks
2. Ackley, D., Hinton, G., & Sejnowski, T. (1985). A learning algorithm for Boltzmann machines. *Cognitive Science*, 9, 147–169.
3. Adams, S. S., Arel, I., Bach, J., Coop, R., Furlan, R., Goertzel, B., . . . Sowa, J. (2012). Mapping the landscape of human-level artificial general intelligence. *AI Magazine*, 33(1), pp. 25-42.
4. AI Image Generator - Copyright Litigation (U.S. District Court for the Northern District of California January 13, 2023).
5. Albus, J. S. (1993). A reference model architecture for intelligent systems design. In P. Antsaklis, & K. Passino (Eds.), *An Introduction to Intelligent and Autonomous Control* (pp. 27-56). Kluwer Academic Publishers.
6. Altman, S. (2023, February 3). OpenAI's Sam Altman talks ChatGPT and how Artificial General Intelligence can 'break capitalism'. (A. Conrad, & K. Cai, Interviewers) Forbes. Retrieved February 16, 2023
7. Anderson, J. R. (1983). *The architecture of cognition*. Harvard: Harvard University Press.
8. *Artificial Intelligence | An Introduction*. (n.d.). Retrieved January 28, 2022, from <https://www.geeksforgeeks.org/>: <https://www.geeksforgeeks.org/artificial-intelligence-an-introduction/>
9. Asimov, I. (1942, March). Roundabout. *Astounding Science Fiction*.
10. Atanasova, T. (2005). *Intelligent computer systems (in Bulgarian)*. Varna: Science and Economy.
11. *Automated driving - levels of driving automation are defined in new SAE International standard J3016*. (2016, March). Retrieved October 14, 2020, from <https://web.archive.org/>:
https://web.archive.org/web/20180701034327/https://cdn.oemoffhighway.com/files/base/acbm/ooh/document/2016/03/automated_driving.pdf
12. Ballard, D. H., & Brown, C. M. (1982). *Computer vision*. Prentice Hall.
13. Beaulieu, S., Frati, L., Miconi, T., Lehman, J., Stanley, K. O., Clune, J., & Cherney, N. (2020). Learning to continually learn. *ECAI 2020*. Santiago de Compostela.
14. Bellamy, R. K., Dey, K., Hind, M., Hoffman, S. C., Houde, S., Kannan, K., . . . Zhang, Y. (2018, October 3). *AI fairness360: An extensible toolkit for detecting, understanding, and mitigating unwanted algorithmic bias*. Retrieved February 23, 2023, from arXiv:1810.01943.
15. Bellman, R. E. (1978). *An introduction to Artificial Intelligence: Can computers think?* Boyd&Fraser Publishing Company.

16. Berg, T., Burg, V., Gombović, A., & Puri, M. (2018, September). *On the rise of the FinTechs—credit scoring using digital footprints*. Retrieved February 23, 2023, from <https://www.fdic.gov/>: <https://www.fdic.gov/analysis/cfr/2018/wp2018/cfr-wp2018-04.pdf>
17. Bhattacharjee, D., Seeley, J., & Seitzman, N. (2017, October 3). *Advanced analytics in hospitality*. Retrieved October 22, 2020, from <https://www.mckinsey.com/>: <https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/advanced-analytics-in-hospitality>
18. Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.
19. Bonawitz, K., Ivanov, V., Kreuter, B., Marcedone, A., McMahan, H. B., Patel, S., . . . Seth, K. (2017). Practical secure aggregation for privacy-preserving machine learning. *ACM SIGSAC Conference on Computer and Communications*, (pp. 1175–1191). Dallas.
20. Bostrom, N. (2014). *Superintelligence: Paths, dangers, strategies*. Oxford: Oxford University Press.
21. Brattberg, E., Csernaton, R., & Rugova, V. (2020, July). *Europe and AI: Leading, lagging behind, or carving its own way?* Retrieved February 26, 2023, from <https://carnegieendowment.org/>: https://carnegieendowment.org/files/BrattbergCsernatonRugova_-_Europe_AI.pdf
22. British Standards Institution. (2016, April). *BS 8611 Robots and robotic devices. Guide to the ethical design and application of robots and robotic systems*. Retrieved February 24, 2023, from <https://standardsdevelopment.bsigroup.com/>: <https://standardsdevelopment.bsigroup.com/projects/9021-05777#/section>
23. Brooks, R. (1990). Elephants don't play chess. *Robotics and Autonomous Systems*, 6(1-2), 3–15.
24. Brooks, R. (2014, November 12). *Artificial intelligence is a tool, not a threat*. Retrieved March 27, 2020, from <http://www.rethinkrobotics.com/>: <https://web.archive.org/web/20141112152108/http://www.rethinkrobotics.com/category/rethinking-robotics/>
25. Bryliuk, D., & Starovoitov, V. (2001). Application of recirculation neural network and principal component analysis for face recognition. *The 2nd International Conference on Neural Networks and Artificial Intelligence*, (pp. 136-142). Minsk.
26. Bryson, J. J., & Winfield, A. (2017). Standardizing ethical design for artificial intelligence and autonomous systems. *Computer*, 50, 116-119.
27. Buchanan, B. G., & Headrick, T. E. (1970). Some speculation about artificial intelligence and legal reasoning. *Stanford Law Review*, 40-62.
28. Busby, M. (2018, April 30). *Revealed: how bookies use AI to keep gamblers hooked*. Retrieved October 23, 2020, from <https://www.theguardian.com/>: <https://www.theguardian.com/technology/2018/apr/30/bookies-using-ai-to-keep-gamblers-hooked-insiders-say>
29. Carpenter, G. A., & Grossberg, S. (1987). ART 2: Self-organization of stable category recognition codes for analog input patterns. *Applied Optics*, 26(23), pp. 4919-4930.
30. Cellan-Jones, R. (2014, December 2). *Stephen Hawking warns artificial intelligence could end mankind*. Retrieved March 26, 2020, from BBC News: <https://www.bbc.com/news/technology-30290540>

31. Chalmers, D. (1995). Facing up to the problem of consciousness. *Journal of Consciousness Studies*, 2(3), pp. 200–219.
32. Chalmers, D. (2010). The singularity: A philosophical analysis. *Journal of Consciousness Studies*, 17, 7–65.
33. Charlevoix common vision for the future of artificial intelligence. (2018, June 9). Retrieved February 26, 2023, from https://www.international.gc.ca/https://www.international.gc.ca/world-monde/assets/pdfs/international_relations-relations_internationales/g7/2018-06-09-artificial-intelligence-artificielle-en.pdf
34. Charniak, E., & McDermott, D. (1985). *Introduction to Artificial Intelligence*. Addison-Wesley.
35. Chui, M., Hall, B., Mayhew, H., Singla, A., & Sukharevsky, A. (2022, December 6). *The state of AI in 2022 - and a half decade in review*. Retrieved February 26, 2023, from <https://www.mckinsey.com/https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-in-2022-and-a-half-decade-in-review/>
36. Clark, J. (2016, July 20). *New Google AI brings automation to customer service*. Retrieved November 1, 2020, from <https://www.bloomberg.com/https://www.bloomberg.com/news/articles/2016-07-20/new-google-ai-services-bring-automation-to-customer-service-iqv2rshg>
37. Clark, J., & Hadfield, G. K. (2019, December 11). *Regulatory markets for AI safety*. Retrieved February 27, 2023, from <https://arxiv.org/https://arxiv.org/pdf/2001.00078.pdf>
38. Clemens Brunner, N. B.-P. (2015, February 10). BNCI Horizon 2020: towards a roadmap for the BCI. *Brain-Computer Interfaces*, 2(1), pp. 1-10. doi:10.1080/2326263X.2015.1008956
39. Cohen, P. R. (1995). *Empirical methods for artificial intelligence*. MIT Press.
40. Culbertson, J. T. (1948). The mechanism for optic nerve conduction and form perception. *Bulletin of Mathematical Biophysics*, 10, 31-40.
41. Culbertson, J. T. (1950). *Consciousness and behavior*. Dubuque, Iowa: Brown.
42. Dastin, J. (2018, October 11). *Amazon scraps secret AI recruiting tool that showed bias against women*. Retrieved February 23, 2023, from <https://www.reuters.com/https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G>
43. Dean, T., & Boddy, M. (1988). An analysis of time-dependent planning. *AAAI-88*, (pp. 49-54). Saint Paul.
44. Dechter, R. (1986). *Learning while searching in constraint-satisfaction-problems*. Retrieved from [www.researchgate.net:https://www.researchgate.net/publication/221605378_Learning_While_Searching_in_Constraint-Satisfaction-Problems](https://www.researchgate.net/https://www.researchgate.net/publication/221605378_Learning_While_Searching_in_Constraint-Satisfaction-Problems)
45. Domingos, P. (2015). *The master algorithm: How the quest for the ultimate learning machine will remake our world*. Basic Books.
46. Dressel, J., & Farid, H. (2018, January 17). The accuracy, fairness, and limits of predicting recidivism. *Science Advances*, 4(1). doi:10.1126/sciadv.aao5580
47. Dreyfus, H. (1972). *What computers can't do*. New York: MIT Press.
48. Eadicicco, L. (2015, January 28). *Bill Gates: Elon Musk Is right, we should all be scared of artificial intelligence wiping out humanity*. Retrieved February 23, 2023, from

<https://www.businessinsider.com/>: <https://www.businessinsider.com/bill-gates-artificial-intelligence-2015-1>

49. Ellett, J. (2017, July 27). *New AI-based tools are transforming social media marketing*. Retrieved October 23, 2020, from <https://www.forbes.com/>: <https://www.forbes.com/sites/johnellett/2017/07/27/new-ai-based-tools-are-transforming-social-media-marketing/#55d66eeb69a2>

50. European Commission. (2021, April 21). *Proposal for a regulation laying down harmonised rules on artificial intelligence*. Retrieved February 15, 2023, from <https://digital-strategy.ec.europa.eu/>: <https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-laying-down-harmonised-rules-artificial-intelligence>

51. European Commission. (2022, August 31). *Artificial intelligence*. Retrieved February 25, 2023, from <https://digital-strategy.ec.europa.eu/>: <https://digital-strategy.ec.europa.eu/en/policies/artificial-intelligence>

52. Evans, W. (2015). Posthuman rights: Dimensions of transhuman worlds. *Teknokultura*, 12(2). doi:10.5209/rev_TK.2015.v12.n2.49072

53. Feigenbaum, E. A., Buchanan, B. G., & Lederberg, J. (1971). On generality and problem solving: A case study using the DENDRAL program. In B. Meltzer, & D. Mitchie (Eds.), *Machine Intelligence 6* (pp. 165-190). Edinburgh University Press.

54. Finn, C., Abbeel, P., & Levine, S. (2017). Model-agnostic meta-learning for fast adaptation of deep networks. *34 th International Conference on Machine Learning*. Sydney.

55. *Five technologies changing agriculture*. (2016, October 7). Retrieved October 12, 2020, from <https://idealog.co.nz/>: <https://idealog.co.nz/tech/2016/10/five-technologies-changing-agriculture>

56. Food and Drug Administration. (2021). *2021: Advancing regulatory science at FDA: Focus areas of regulatory science*. Retrieved January 27, 2023, from <https://www.fda.gov/>: <https://www.fda.gov/media/145001/download>

57. Ford, M. (2018). *Architects of intelligence*. Packt.

58. Ford, M., & Colvin, G. (2015, September 6). Will robots create more jobs than they destroy? *The Guardian*.

59. Fox, M. (1986). Industrial applications of artificial Intelligence. *Robotics*, 2(4), 301–311. doi:10.1016/0167-8493(86)90003-3

60. Franklin, S., & Patterson, F. G. (2006). The LIDA architecture: Adding new modes of learning to an intelligent, autonomous, software agent. *IDPT-2006 Proceedings (Integrated Design and Process Technology): Society for Design and Process Science*.

61. Freeman, J. A., & Skapura, D. M. (1991). *Neural networks: Algorithms, applications, and programming techniques*. Addison-Wesley.

62. Frey, C. B., & Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerisation? *Technological forecasting and social change*, 114, 254–280.

63. Fukushima, K. (1975). Cognitron: A self-organizing multilayered neural network. *Biological Cybernetics*, 20.

64. Fukushima, K. (1980). Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, 36(4), 193-202. doi:10.1007/bf00344251

65. Galushkin, A. I. (2010). *Neural networks basics (in Russian)*. Moskow: Hot line telekom.
66. Geist, E. M. (2015, July 30). *Is artificial intelligence really an existential threat to humanity?* Retrieved February 23, 2023, from <https://web.archive.org/https://web.archive.org/web/20151030054330/http://thebulletin.org/artificial-intelligence-really-existential-threat-humanity8577>
67. George, B., & Carmichael, G. (2015). *Artificial intelligence simplified: Understanding basic concepts*. Bettendorf: CSTrends LLP.
68. Gibbs, S. (2014, October 27). *Elon Musk: artificial intelligence is our biggest existential threat*. Retrieved February 23, 2023, from <https://www.theguardian.com/https://www.theguardian.com/technology/2014/oct/27/elon-musk-artificial-intelligence-ai-biggest-existential-threat>
69. Gil, Y., King, R., & Kitano, H. (2020, February). *Posing an AI scientist grand challenge: Artificial intelligence systems capable of Nobel-quality discoveries*. Retrieved February 16, 2023, from https://www.turing.ac.uk/https://www.turing.ac.uk/sites/default/files/2021-02/summary_of_discussion_workshop_2020_ai_scientist_grand_challenge_clean.pdf
70. Girgin, S. (2019, May 22). *Decision tree regression in 6 steps with Python*. Retrieved February 1, 2023, from <https://medium.com/https://medium.com/pursuitnotes/decision-tree-regression-in-6-steps-with-python-1a1c5aa2ee16>
71. Goertzel, B. (2014, December 30). Artificial general intelligence: Concept, state of the art, and future prospects. *Journal of Artificial General Intelligence*, 5(1), 1-48. doi:10.2478/jagi-2014-0001
72. Goertzel, B., & Bugaj, S. V. (2009). AGI preschool. *Second Conference on Artificial General Intelligence (AGI09)*. Atlantis Press.
73. Goertzel, B., Lian, R., Arel, I., de Garis, H., & Chen, S. (2010, December). A world survey of artificial brain projects, Part II: Biologically inspired cognitive architectures. *Neurocomputing*, 74(1-3), pp. 30–49. doi:10.1016/j.neucom.2010.08.01
74. Good, I. J. (1965). Speculations concerning the first ultraintelligent machine. *Advances in Computers*, 6, 31-83.
75. Goodfellow, I. J., Shlens, J., & Szegedy, C. (2015, March 20). *Explaining and harnessing adversarial examples*. Retrieved January 19, 2023, from <https://arxiv.org/https://arxiv.org/pdf/1412.6572.pdf>
76. GPAI Council. (2022, November 22). *GPAI 2022 ministers' declaration*. Retrieved February 24, 2023, from <https://www.gpai.ai/https://www.gpai.ai/events/tokyo-2022/ministerial-declaration/GPAIMinistersDeclaration2022.pdf>
77. Grace, K., Salvatier, J., Dafoe, A., Zhang, B., & Evans, O. (2017). When will AI exceed human performance? Evidence from AI experts. arXiv:1705.08807. Retrieved March 1, 2023
78. Grossberg, S. (1969). Some networks that can learn, remember, and reproduce any number of complicated space-time patterns, I. *Journal of Mathematics and Mechanics*, 19.
79. Grossberg, S. (1970). Some networks that can learn, remember, and reproduce any number of complicated space-time patterns, II. *Studies in Applied Mathematics*, 49.

80. Grossberg, S. (1976). Adaptive pattern classification and universal recoding: I. Parallel development and coding of neural feature detectors. *Biological Cybernetics*, 23, 187-202.
81. Grosz, B. (2012, December 12). What question would Turing pose today? 33(4), pp. 73-81. doi:10.1609/aimag.v33i4.2441
82. Gulf News. (2018, July 28). *UAE and India sign agreement on Artificial Intelligence*. Retrieved February 26, 2023, from <https://gulfnews.com/https://gulfnews.com/uae/government/uae-and-india-sign-agreement-on-artificial-intelligence-1.2258074#>
83. Haenlein, M., & Kaplan, A. (2019). A brief history of artificial intelligence: On the past, present, and future of Artificial Intelligence. *California Management Review*, 61(4), 5–14. doi:10.1177/0008125619864925
84. Harnad, S. (2001). What's wrong and right about searle's chinese room argument? In M. Bishop, & J. Preston (Eds.), *Essays on Searle's Chinese Room Argument*. Oxford University Press.
85. Haugeland, J. (1985). *Artificial Intelligence: The very idea*. Cambridge, Massachusetts: MIT Press.
86. Haven, J., & Boyd, D. (2020, November 9). *Philanthropy's techno-solutionism problem*. Retrieved January 20, 2023, from <https://knightfoundation.org/https://knightfoundation.org/wp-content/uploads/2020/11/KF-Kettering-Philanthropy-Techno-Solution-Problem.pdf>
87. Hay, N., Russell, S. J., Shimony, S. E., & Tolpin, D. (2012). Selecting computations: Theory and applications. *UAI-12*.
88. Haykin, S. (1999). *Neural networks: A comprehensive foundation* (2nd ed.). Delhi: Pearson Education, Inc.
89. He, H., Bai, Y., Garcia, E. A., & Li, S. (2008). ADASYN: Adaptive synthetic sampling approach for imbalanced learning. *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, (pp. 1322 - 1328). Hong Kong.
90. Hebb, D. O. (1949). *The organization of behavior*. New York: Wiley.
91. Hecht-Nielsen, R. (1987). Counter-propagation networks. *First International Conference on Neural Networks, Volume II*.
92. Hernandez-Orallo, J., & Dowe, D. L. (2010). Measuring universal intelligence: Towards an anytime intelligence test. *Artificial Intelligence*, 174(18), 1508-1539.
93. Hibbard, B. (2002). *Super-intelligent machines*. Kluwer Academic/Plenum Publishers.
94. Hinton, G. E. (2007). Learning multiple layers of representation. *Trends in Cognitive Science*, 11(10), 428-434. doi:10.1016/j.tics.2007.09.004
95. Hinton, G. E., & McClelland, J. L. (1988). Learning representations by recirculation. *IEEE Conference on Neural Information Processing Systems*.
96. Hinton, G. E., & Sejnowski, T. J. (1983). Optimal perceptual inference. *CVPR*, (pp. 448-453). Washington DC.
97. Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, 9(8), 1735-1780.

98. Hooke, K. (2020, July 23). *A deep dive into the transformer architecture – the development of transformer models*. Retrieved February 7, 2023, from <https://dzone.com/articles/a-deep-dive-into-the-transformer-architecture-the>
99. Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *PNAS*, 79, pp. 2554-2558.
100. Hopfield, J. J., & Tank, D. W. (1985). "Neural" computation of decisions in optimization problems. *Biological Cybernetics*, 55, 141-146.
101. Horst, S. (2015, October 16). *The computational theory of mind*. Retrieved from The Stanford Encyclopedia of Philosophy: <https://plato.stanford.edu/entries/computational-mind/>
102. Hristov, A., Nisheva, M., & Dimov, D. (2018). An introduction into convolutional neural networks. *Automatica and Informatics*(1), 27-38.
103. IEEE Standards Association. (2016, December). *Ethically aligned design: A vision for prioritizing human well-being with autonomous and intelligent systems*. Retrieved February 24, 2023, from <https://ieeexplore.ieee.org/https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9398613>
104. ISO. (2022, August). *ISO/IEC TR 24368:2022 information technology - artificial intelligence - overview of ethical and societal concerns*. Retrieved February 25, 2023, from <https://www.iso.org/https://www.iso.org/standard/78507.html?browse=tc>
105. Jacobs, R. A. (1988). Increased rates of convergence through learning rate adaptation. *Neural Networks*, 1.
106. Jain, L., & Fanelli, A. M. (2000). *Recent advances in artificial neural networks: Design and application*. CRC Press LLC.
107. James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning*. Springer.
108. Jankovic, J. (2020, December 1). <https://srinstitute.utoronto.ca/>. Retrieved February 27, 2023, from Schwartz Reisman Institute and AI Global to develop global certification for trustworthy AI: <https://srinstitute.utoronto.ca/news/ai-global-certification-partnership>
109. Jefferson, G. (1949). The mind of mechanical man: The Lister Oration delivered at the Royal College of Surgeons in England. *British Medical Journal*, 1(25), pp. 1105-1121.
110. Jobin, A., Ienca, M., & Vayena, E. (2019, September 2). Artificial Intelligence: The global landscape of ethics guidelines. *Nat Mach Intell*, 1, pp. 389-399. doi:10.1038/s42256-019-0088-2
111. John Wiley & Sons, Inc. (2001). *Kalman filtering and neural networks*. (S. Haykin, Editor) Retrieved February 1, 2020, from https://onlinelibrary.wiley.com/https://onlinelibrary.wiley.com/doi/pdf/10.1002/0471221546.fmatter_indsb
112. Johnson, C. Y. (2019, October 24). *Racial bias in a medical algorithm favors white patients over sicker black patients*. Retrieved February 23, 2023, from <https://www.washingtonpost.com/https://www.washingtonpost.com/health/2019/10/24/racial-bias-medical-algorithm-favors-white-patients-over-sicker-black-patients/>
113. Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82, 34-45.

114. Kaplan, A., & Haenlein, M. (2019). Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. *Business Horizons*, 62, 15–25. doi:10.1016/j.bushor.2018.08.004
115. Kaplan, J. (2017, March 3). *AI's PR problem*. Retrieved November 18, 2020, from <https://www.technologyreview.com/>
<https://www.technologyreview.com/2017/03/03/153435/ais-pr-problem-2/>
116. Katz, Y. (2012, November 1). *Noam Chomsky on where artificial intelligence went wrong*. Retrieved February 13, 2020, from www.theatlantic.com.
117. Kelly, S. M. (2023, January 26). *ChatGPT passes exams from law and business schools*. Retrieved January 31, 2023, from <https://cnn.com/>:
<https://edition.cnn.com/2023/01/26/tech/chatgpt-passes-exams/index.html>
118. Kenessey, Z. (2012). *The primary, secondary, tertiary and quaternary sectors of the Economy*. The Review of Income and Wealth, US. Federal Reserve Board. Retrieved January 27, 2023, from <http://www.roiw.org/1987/359.pdf>
119. Khemali, C., Doshi, J., Duseja, J., Shan, K., Udmale, S., & Sambhe, V. (2019). Solving Rubik's cube using graph theory: ICCI-2017. In C. Khemali, J. Doshi, J. Duseja, K. Shan, S. Udmale, & V. Sambhe, *Computational intelligence: Theories, applications and future directions-volume 1* (pp. 301-317). Springer Nature Singapore Pte Ltd.
120. Kitano, H. (Spring 2016). Artificial intelligence to win the Nobel Prize and beyond: Creating the engine for scientific discovery. *AI Magazine*, 37(1).
121. Klaua, D. (1965). Über einen ansatz zur mehrwertigen mengenlehre. *Monatsb. Deutsch. Akad. Wiss. Berlin* 7, 859–876.
122. Knuth, D. (1998). Sorting and Searching. *The Art of Computer Programming*, 3.
123. Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, 43, 59–69.
124. Kohonen, T. (1988). Learning vector quantization. *Neural Networks*, 1.
125. Kohonen, T. (1988). *Self-organization and associative memory* (2nd ed.). Springer.
126. Kohonen, T. (1997). *Self-organizing maps*. Berlin: Springer-Verlag.
127. Kosko, B. (1987). Adaptive bidirectional associative memories. *Applied Optics*, 26.
128. Kosko, B. (1988). Bidirectional associative memories. *IEEE Transactions on Systems, Man, and Cybernetics*, 18(1), 49-60.
129. Kurzweil, R. (1990). *The age of intelligent machines*. MIT Press.
130. Kurzweil, R. (2005). *The singularity is near*. Viking.
131. Laird, J. E. (2012). *The Soar cognitive architecture*. MIT Press.
132. Laird, J. E., Wray, R., Marinier, R., & Langley, P. (2009). Claims and challenges in evaluating human-level intelligent systems. *Second Conference on Artificial General Intelligence*, (pp. 91-96).
133. Lander, E., Baylis, F., Zhang, F., Charpentier, E., Berg, P., Bourgain, C., . . . Winnacker, E.-L. (2019, March 14). Adopt a moratorium on heritable genome editing. *Nature*, 567, 165-168. doi:10.1038/d41586-019-00726-5
134. Langley, P. (2011). The changing science of machine learning. *Machine Learning*, 82(3), pp. 275–279. doi:10.1007/s10994-011-5242-y

135. Laricchia, F. (2022, November 23). *Video surveillance camera market size worldwide from 2019 to 2026*. Retrieved February 21, 2023, from <https://www.statista.com/https://www.statista.com/statistics/477917/video-surveillance-equipment-market-worldwide/>
136. Lebiere, C., & Anderson, J. R. (1993). A connectionist Implementation of the ACT-R production system. *Proceedings of the Fifteenth Annual Conference of the Cognitive Science Society* (pp. 635-640). Mahwah, NJ: Lawrence Erlbaum Associates.
137. Lee, T. B. (2015, July 29). *Will artificial intelligence destroy humanity? Here are 5 reasons not to worry*. Retrieved February 23, 2023, from <https://www.vox.com/https://www.vox.com/2014/8/22/6043635/5-reasons-we-shouldnt-worry-about-super-intelligent-computers-taking>
138. Legg, S., & Veness, J. (2013). An approximation of the universal intelligence measure. In *Algorithmic Probability and Friends. Bayesian Prediction and Artificial Intelligence* (pp. 236–249). Springer.
139. Lenat, D. B., & Guha, R. V. (1990). *Building large knowledge-based systems: Representation and inference in the CYC project*. Addison-Wesley.
140. Li, T.-M., Gharbi, M., Adams, A., Durand, F., & Ragan-Kelley, J. (2018, August). Differentiable programming for image processing and deep learning in Halide. *ACM Transactions on Graphics*, 37(4).
141. Lippmann, R. P. (1987, April). An introduction to computing with neural nets. *IEEE ASSP Magazine*.
142. Littman, M. L., Ajunwa, I., Berger, G., Boutilier, C., Currie, M., Doshi-Velez, F., . . . Walsh, T. (2021, September). *Gathering strength, gathering storms: The one hundred year study on Artificial Intelligence (AI100) 2021 study panel report*. Retrieved January 22, 2023, from https://ai100.stanford.edu/https://ai100.stanford.edu/sites/g/files/sbiybj18871/files/media/file/AI100Report_MT_10.pdf
143. Lum, K., & Isaac, W. (2016, October). *To predict and serve?* Retrieved February 23, 2023, from [www.significancemagazine.com:https://rss.onlinelibrary.wiley.com/doi/epdf/10.1111/j.1740-9713.2016.00960.x](https://rss.onlinelibrary.wiley.com/doi/epdf/10.1111/j.1740-9713.2016.00960.x)
144. Marinova, N. (2014). *Artificial intelligence systems*. Svishtov: Tsenov Publishing House.
145. Marwala, T., & Hurwitz, E. (2017). *Artificial intelligence and economic theory: Skynet in the market*. London: Springer.
146. McCarthy. (1958). Programs with common sense. *Symposium on Mechanisation of Thought Processes*, 1, pp. 77-84.
147. McCarthy, J. (1980, April). Circumscription – a form of non-monotonic reasoning. *Artificial Intelligence*, 13, 27–39. doi:10.1016/0004-3702(80)90011-9
148. McCarthy, J., & Hayes, P. J. (1969). Some philosophical problems from the standpoint of Artificial Intelligence. In B. Meltzer, D. Michie, & M. Swann (Ed.), *Machine Intelligence 4* (pp. 463-502). Edinburgh University Press.
149. McCarthy, J., Minsky, M., Rochester, N., & Shannon, C. E. (1955). *Proposal for the Dartmouth summer research project on artificial intelligence*. Dartmouth College.

150. McClelland, J. L., Rumelhart, D. E., & Group, P. R. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 2). Cambridge, Mass: MIT Press.
151. McCorduck, P. (2004). *Machines who think* (2nd ed.). Natick, MA: A. K. Peters, Ltd.
152. McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5, 115-133.
153. Meehan, J. R. (1977). Tale-spin, an interactive program that writes stories. *Proceedings of the 5th International Joint Conference on Artificial Intelligence - Volume 1, IJCAI'77* (pp. 91-98). San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.
154. Mehr, H. (2017, August). *Artificial intelligence for citizen services and government*. Retrieved October 26, 2020, from https://ash.harvard.edu/https://ash.harvard.edu/files/ash/files/artificial_intelligence_for_citizen_services.pdf
155. Minai, A. A., & Williams, R. D. (1990). Acceleration of back-propagation through learning rate and momentum adaptation. *International Joint Conference on Neural Networks, 1*.
156. Ministri of Transport and Communications. (October 2020 r.). *Concept for the development of artificial intelligence in Bulgaria by 2030: Artificial intelligence for smart growth and a prosperous democratic society*. Изтеглено на 25 Февруари 2023 г. от <https://www.mtc.government.bg/https://www.mtc.government.bg/sites/default/files/koncepciyazarazvitenaiivbulgariyado2030.pdf>
157. Mitchell, M. (2019). *Artificial intelligence: A guide for thinking humans*. New York: Farrar, Straus and Giroux.
158. Mittelstadt, B. (2019). Principles alone cannot guarantee ethical AI. *Nature Machine Intelligence, 1*, 501-507.
159. Moor, J. M. (2009). Four kinds of ethical robots. *Philosophy Now*.
160. Moravec, H. (1988). *Mind children: The future of robot and human intelligence*. Harvard University Press.
161. Moravec, H. (1997, October). (N. Online, Interviewer) Retrieved January 10, 2020, from <https://www.pbs.org/wgbh/nova/robots/moravec.html>
162. Moravec, H. P. (2000). *Robot: Mere machine to transcendent mind*. Oxford University Press.
163. Muehlhauser, L. (2013, July 31). *AI risk and the security mindset*. Retrieved February 23, 2023, from <https://intelligence.org/https://intelligence.org/2013/07/31/ai-risk-and-the-security-mindset/>
164. Muehlhauser, L. (2013, August 11). *What is AGI?* Retrieved March 11, 2020, from Machine Intelligence Research Institute: <https://web.archive.org/web/20140425115445/http://intelligence.org/2013/08/11/what-is-agi/>
165. Murray, P. (2012, March 21). *Amazon goes robotic, acquires Kiva Systems, makers of warehouse robot*. Retrieved October 23, 2020, from <https://singularityhub.com/https://singularityhub.com/2012/03/21/amazon-goes-robotic-acquires-kiva-systems-makers-of-the-warehouse-robot/>

166. Neikov, P., & Toshkova, V. (17 Януари 2023 г.). *2023 could be a turning point for technology regulations (in Bulgarian)*. Изтеглено на 25 Февруари 2023 г. от <https://www.investor.bg/>: <https://www.investor.bg/a/566-novini-i-analizi/367218-2023-g-mozhe-da-se-okazhe-povratna-za-tehnologichnite-regulatsii>
167. NeuralWare, Inc. (1991). *Neural computing: Neural-works Professional II/Plus ANN development software*. Pittsburg: NeuralWare, Inc.
168. *New sensors make for soft and sensitive robotic fingers*. (2020, December 15). Retrieved January 18, 2023, from <https://www.processonline.com>: <https://www.processonline.com.au/content/factory-automation/news/new-sensors-make-for-soft-and-sensitive-robotic-fingers-1590420400>
169. Newell, A., & Simon, H. A. (1976). Computer science as empirical inquiry: Symbols and search. *Communications of the ACM*, 19.
170. Nicas, J. (2020, December 8). *Why can't the social networks stop fake accounts?* Retrieved December 16, 2022, from <https://www.nytimes.com/>: <https://www.nytimes.com/2020/12/08/technology/why-cant-the-social-networks-stop-fake-accounts.html>
171. Nilsson, N. J. (1998). *Artificial Intelligence: A new synthesis*. Morgan Kaufmann.
172. Nilsson, N. J. (2005). Human-level artificial intelligence? Be serious! *AI Magazine*, 26(4), pp. 68-75.
173. Nilsson, N. J. (2010). *The quest for Artificial Intelligence: A history of ideas and achievements*. California: Cambridge University Press.
174. Nordic co-operation. (2018, May 14). *AI in the Nordic-Baltic region*. Retrieved February 26, 2023, from <https://www.norden.org/>: <https://www.norden.org/en/declaration/ai-nordic-baltic-region>
175. O'Neil, C. (2017). *Weapons of math destruction: How big data Increases Inequality and threatens democracy*. Crown: Broadway Books.
176. OECD. (2021, June). *State of implementation of the OECD AI principles: Insights from national AI policies*. Retrieved February 25, 2023, from <https://www.oecd-ilibrary.org/>: https://www.oecd-ilibrary.org/science-and-technology/state-of-implementation-of-the-oecd-ai-principles_1cd40c44-en
177. OECD.AI. (2021). *National AI policies & strategies*. Retrieved February 25, 2023, from <https://oecd.ai/>: <https://oecd.ai/en/dashboards/overview>
178. Olewitz, C. (2016, March 23). *A Japanese A.I. program just wrote a short novel, and it almost won a literary prize*. Retrieved November 16, 2020, from <https://www.digitaltrends.com/>: <https://www.digitaltrends.com/cool-tech/japanese-ai-writes-novel-passes-first-round-nationnl-literary-prize/>
179. Omohundro, S. M. (2008, February). The basic AI drives. *AGI*, 171, 483-492.
180. Open Community for Ethics in Autonomous and Intelligent Systems. (2022). *IEEE P7000™ Projects*. Retrieved February 24, 2023, from <https://ethicsstandards.org/p7000/>: <https://ethicsstandards.org/p7000/>
181. Oxford Insights. (n.d.). *Government AI readiness index 2020*. Retrieved February 26, 2023, from <https://static1.squarespace.com/>: <https://static1.squarespace.com/static/58b2e92c1e5b6c828058484e/t/5f7747f29ca3c20ecb598f7c/1601653137399/AI+Readiness+Report.pdf>

182. Pao, Y.-H. (1989). *Adaptive pattern recognition and neural networks*. Boston, MA: Addison-Wesley Longman Publishing Co., Inc.
183. Parker, D. B. (1987). Optimal algorithms for adaptive networks: Second order back propagation, second order direct propagation and second order hebbian learning. *ICNN, II*.
184. Parliament of Canada. (2020, November 17). *BILL C-11*. Retrieved March 2, 2023, from <https://parl.ca/>: <https://parl.ca/DocumentViewer/en/43-2/bill/C-11/first-reading>
185. Parzen, E. (1962). On the estimation of a probability density function and the mode. *Annals of Mathematical Statistics*, 33, 1065-1076.
186. Pfeifer, R., & Bongard, J. (2007). *How the body shapes the way we think: A new view of intelligence*. MIT press.
187. Picard, R. W. (1995). *Affective computing*. Cambridge: M.I.T Media Laboratory Perceptual Computing Section Technical Report No. 321. Retrieved from <https://affect.media.mit.edu/pdfs/95.picard.pdf>
188. Pitts, W., & McCulloch, W. S. (1947). How we know universals: The perception of auditory and visual forms. *Bulletin of Mathematical Biophysics*, 9, 127-147.
189. Pogled.info. (11 Януари 2023 г.). *Първите разпоредби на Китай за динфейк технологията вече са в сила*. Изтеглено на 25 Февруари 2023 г. от <https://pogled.info/>: <https://pogled.info/svetoven/pogled-kitai/parvite-razporedbi-na-kitai-za-dipfeik-tehnologiyata-veche-sa-v-sila.151042>
190. Poole, D. L., & Mackworth, A. K. (2017). *Artificial Intelligence: Foundations of computational agents* (2nd ed.). Cambridge: Cambridge University Printing House. doi:10.1017/9781108164085
191. Poole, D., Mackworth, A. K., & Goebel, R. (1998). *Computational intelligence: A logical approach*. Oxford University Press.
192. Puskorius, G. V., & Feldkamp, L. A. (1991). Decoupled extended Kalman filter training of feedforward layered networks. *International Joint Conference of Neural Networks*, 1, pp. 771–777. Seattle.
193. Rabuñal, J. R., & Dorado, J. (2006). *Artificial neural networks in real-life applications*. Idea Group Publishing.
194. Rao, D. A., & Verweij, G. (2017). Sizing the prize: What’s the real value of AI for your business and how can you capitalise? *PwC*.
195. Raelison, M., Boissin, E., Borst, G., & Neys, W. D. (2021, April 1). From slow to fast logic: The development of logical intuitions. *Thinking & Reasoning*, 599-622. doi:10.1080/13546783.2021.1885488
196. Reddy, R. (1988). Foundations and grand challenges of Artificial Intelligence. *AI Magazine*, 9(4), 9-21.
197. Reisman, D., Schultz, J., Crawford, K., & Whittaker, M. (2018, April). *Algorithmic impact assessments: A practical framework for public agency accountability*. Retrieved February 27, 2023, from <https://ainowinstitute.org/>: <https://ainowinstitute.org/aiareport2018.pdf>
198. Reiter, R. (1980). A logic for default reasoning. *Artificial Intelligence*, 13, 81-132.
199. RenAIssance Foundation. (2020, February 28). *Rome Call for AI ethics*. Retrieved February 25, 2023, from <https://www.romecall.org/>: https://www.romecall.org/wp-content/uploads/2022/03/RomeCall_Paper_web.pdf

200. Reuters Events. (2023, February 16). The evolution of automotive technology. Retrieved March 2, 2023, from <https://events.reutersevents.com/automotive/technology>
201. Revell, T. (2017, May 31). *AI will be able to beat us at everything by 2060, say experts*. Retrieved June 2, 2020, from <https://www.newscientist.com:https://www.newscientist.com/article/2133188-ai-will-be-able-to-beat-us-at-everything-by-2060-say-experts/>
202. Rich, E., & Knight, K. (1991). *Artificial Intelligence* (2nd ed.). McGraw-Hil.
203. Roberts, J. (2016). Thinking machines: The search for Artificial Intelligence. *Distillations*, 2(2), pp. 14-23. Retrieved from <https://web.archive.org>.
204. Rosenblatt, F. (1957). *The perceptron: A perceiving and recognizing automation*. Report 85-460-1, Project PARA, Cornell Aeronautical Laboratory.
205. Rosenblatt, F. (1962). *Principles of Neurodynamics: Perceptrons and the theory of brain mechanisms*. Washington, DC: Spartan Books.
206. Rubin, C. (2003). Artificial intelligence and human nature . *The New Atlantis*, 1, 88–100.
207. Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986b). Learning representations by back-propagating errors. *Nature*, 323, 533–536. doi:10.1038/323533a0
208. Russel, S., & Norvig, P. (2016). *Artificial Intelligence: A modern approach* (3rd ed.). Essex: Pearson Education Limited.
209. Russel, S., & Norvig, P. (2021). *Artificial Intelligence: A modern approach*. (4th). Hoboken, NJ, USA: Pearson Education Limited. Retrieved February 19, 2023, from <https://archive.org/details/artificial-intelligence-a-modern-approach-4th-edition/page/n11/mode/2up>
210. Russell, S. J. (2019). *Human compatible: Artificial intelligence and the problem of control*. New York: Viking.
211. Russell, S. J., & staff. (2022, April). *CHAI 2022 progress report*. Retrieved February 18, 2023, from <https://humancompatible.ai/>: <https://humancompatible.ai/progress-report/>
212. Sagan, C. (1977). *The dragons of eden*. Random House.
213. Samuel, A. (1959). Some studies in machine learning using the game of checkers. *IBM Journal of Research and Development*, 3(3), pp. 210-229. doi:10.1147/rd.33.0210
214. Schank, R. (1973). Conceptualizations underlying natural language. In R. Schank, & K. Colby (Eds.), *Computer Models of Thought and Language*. San Francisco: W.H. Freeman.
215. Schmidhuber, J. (2015, January). Deep learning in neural networks: An overview. *Neural Networks*, 61, 85-117. doi:10.1016/j.neunet.2014.09.003
216. Schmidt, E., Catz, S., SteveChien, MignonClyburn, Darby, C., Ford, K., . . . Moore, A. (2021). *The national security commission on artificial intelligence: Final report*. Retrieved February 26, 2023, from <https://www.nsc.gov/>: <https://www.nsc.gov/wp-content/uploads/2021/03/Full-Report-Digital-1.pdf>
217. Schulz, H., & Behnke, S. (2012, May 17). Deep learning: Layer-wise learning of feature hierarchies. *KI – Künstliche Intelligenz*, 26, 357-363.
218. Schwartz, E. (1990). *Computational neuroscience*. Cambridge, Massachusetts: MIT Press.

219. Schwartz, J. (1987). Limits of artificial intelligence. In S. Shapiro, & D. Eckroth (Eds.), *Encyclopedia of artificial intelligence* (Vol. 1, pp. 488-503). New York: John Wiley and Sons, Inc.
220. Searle, J. (1980). Minds, brains and programs. *Behavioral and Brain Sciences*, 3(3), pp. 417–457. doi:10.1017/S0140525X00005756
221. Searle, J. (1999). *Mind, language and society*. New York, NY: Basic Books.
222. Sears, A. (2018, April 14). *The role of artificial intelligence in the classroom*. Retrieved October 5, 2020, from <https://elearningindustry.com:https://elearningindustry.com/artificial-intelligence-in-the-classroom-role>
223. Shannon, C. E., & Elwood, C. (1948, july). A mathematical theory of communication. *Bell System Technical Journal*, 27(3), pp. 379-423. doi:10.1002/j.1538-7305.1948.tb01338.x
224. Shapiro, S. C. (1992). AI-Complete Tasks. In S. C. Shapiro (Ed.), *Encyclopedia of Artificial Intelligence* (Second ed., pp. 54–57). New York: John Wiley.
225. Siddique, N., & Adeli, H. (2013). *Computational intelligence: Synergies of fuzzy logic, neural networks and evolutionary computing*. John Wiley & Sons.
226. Simon, H., & Newell, A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice Hall.
227. Singhal, S., & Wu, L. (1989). Training multilayer perceptrons with the extended Kalman algorithm. *Advances in Neural Information Processing Systems*, 1, 133–140.
228. Slyusar, V. (2019, July). *Artificial intelligence as the basis of future control networks*. Preprint. doi:10.13140/RG.2.2.30247.50087
229. Specht, D. F. (1988). Probabilistic neural networks for classification, mapping or associative memory. *ICNN-88*.
230. Specht, D. F. (1990, November). Probabilistic neural networks. *Neural Networks*.
231. Taylor, B. J. (2006). *Methods and procedures for the verification and validation of artificial neural networks*. Fairmont: Springer.
232. Todorov, G. (2005). *Artificial intelligence (in Bulgarian)*. Veliko Tarnovo: Faber.
233. Tomayko, J. E. (2003, April). *The story of self-repairing flight control systems*. (C. Gelzer, Ed.) Retrieved October 20, 2020, from https://crgis.ndc.nasa.gov/:https://crgis.ndc.nasa.gov/crgis/images/c/c9/88798main_srfcs.pdf
234. Tomov, H. (19 January 2023 r.). AI will soon write news in Bulgarian especially for you, but it may be fake (in Bulgarian). (D. Nikolov, Интервюиращ) Блумбърг България.
235. Tononi, G., Boly, M., Massimini, M., & Koch, C. (2016). Integrated information theory: From consciousness to Its physical substrate. *Nature Reviews Neuroscience*, 17, 450-461. Retrieved March 2, 2023, from <https://www.nature.com/articles/nrn-2016-44>
236. Turing, A. (1950). Computing machinery and intelligence. *Mind*, 433-460.
237. Turing, A. M. (1951). Intelligent machinery, a heretical theory. *reprinted Philosophia Mathematica (1996)*, 4(3), 256–260. doi:10.1093/philmat/4.3.256
238. UNESCO. (2021, November 23). *Recommendation on the ethics of artificial intelligence*. Retrieved February 25, 2023, from <https://unesco.org/:https://unesdoc.unesco.org/ark:/48223/pf0000381137>
239. UNESCO. (2021). *UNESCO science report: The race against time for smarter development*. (S. Schneegans, T. Straza, & J. Lewis, Eds.) Paris: UNESCO Publishing.

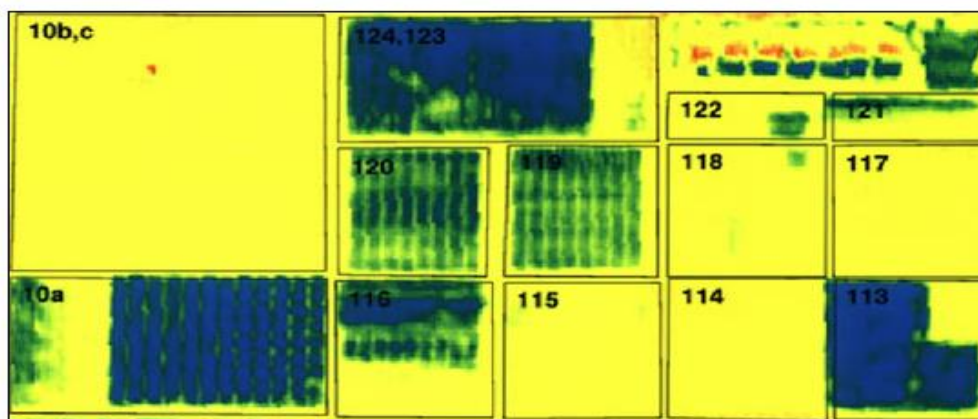
240. UNESCO. (2021). *UNESCO Science Report: The Race Against Time for Smarter Development*. (S. Schneegans, T. Straza, & J. Lewis, Eds.) Paris: UNESCO Publishing. Retrieved February 24, 2023, from UNESCO science report.
241. United Nations Office for Disarmament Affairs. (1980, October 10). *The convention on certain conventional weapons*. Retrieved February 21, 2023, from <https://www.un.org/>: <https://geneva-s3.unoda.org/static-unoda-site/pages/templates/the-convention-on-certain-conventional-weapons/CONVENTION.pdf>
242. Vasquez, H. (2023, February 11). *ChatGPT was close to passing the US doctor's license exam*. Retrieved February 11, 2023, from <https://www.aroged.com/>: <https://www.aroged.com/2023/02/11/chatgpt-was-close-to-passing-the-us-doctors-license-exam/#:~:text=OpenAI%E2%80%99s%20ChatGPT%20Large%20Language%20Model%20%28LLM%29%20algorithm%20almost,only%20on%20searching%20for%20information%20on%20the%20Internet.>
243. Villasenor, J. (2019, January 3). *Artificial intelligence and bias: Four key challenges*. Retrieved March 26, 2020, from Techtank: <https://www.brookings.edu/blog/techtank/2019/01/03/artificial-intelligence-and-bias-four-key-challenges/>
244. Vincent, J. (2023, February 15). *Microsoft's Bing is an emotionally manipulative liar, and people love it*. Retrieved February 16, 2023, from <https://www.theverge.com/>: <https://www.theverge.com/2023/2/15/23599072/microsoft-ai-bing-personality-conversations-spy-employees-webcams>
245. Vinge, V. (1993). The coming technological singularity: How to survive in the post-human era. *VISION-21 Symposium*. NASA Lewis Research Center and the Ohio Aerospace Institute.
246. Von Neuman, J. (1951). *The general and logical theory of automata*. New York: Wiley.
247. Waldrop, M. (1987). A question of responsibility. *AI Magazine*, 8(1), 28-39. doi:10.1609/aimag.v8i1.572
248. Wallach, W. (2010). *Moral machines*. Oxford University Press.
249. Wallach, W., & Allen, C. (2009). *Moral machines: Teaching robots right from wrong*. Oxford University Press, Inc.
250. Wang, P. (2006). *Rigid flexibility: The logic of intelligence*. Springer.
251. Weiss, G. (2013). *Multiagent systems* (2 ed.). Cambridge, MA: The MIT Press.
252. Weizenbaum, J. (1976). *Computer power and human reason*. New York: W. H. Freeman.
253. Widrow, B., & Hoff, M. E. (1960). Adaptive switching circuits. *IRE WESCON Convention Record*, (pp. 96-104).
254. Wiener, N. (1948). *Cybernetics: Or control and communication in the animal and the machine*. Paris: MIT Press.
255. Wilson, D. H. (2011). *Robocalypse*. New York: Doubleday.
256. Winograd, T. (1972). Understanding natural language. *Cognitive Psychology*, 3(1), pp. 1-191.
257. Winston, P. H. (1992). *Artificial Intelligence* (3rd ed.). Reading: Addison-Wesley.

258. WIPO. (2019). *WIPO technology trends 2019: Artificial intelligence*. Geneva: World Intellectual Property Organization.
259. Wissner-Gross, A. D., & Freer, C. (2013, April 19). Causal entropic forces. *110*, 168702. doi:10.1103/PhysRevLett.110.168702
260. World Economic Forum. (2019). *Guidelines for AI procurement*. Retrieved February 25, 2023, from https://www3.weforum.org/https://www3.weforum.org/docs/WEF_Guidelines_for_AI_Procurement.pdf
261. Wozniak, S., & Moon, P. (2007, July 19). Three minutes with Steve Wozniak. PC World.
262. *Yann LeCun sparks a debate on AGI vs human-level AI*. (2022, January 27). (Akashdeep Arul) Retrieved February 24, 2023, from <https://analyticsindiamag.com/https://analyticsindiamag.com/yann-lecun-sparks-a-debate-on-agi-vs-human-level-ai/>
263. Yudkowsky, E. (2008). Artificial intelligence as a positive and negative factor in global risk. In N. Bostrom, & M. Cirkovich (Eds.), *Global Catastrophic Risk*. Oxford University Press.
264. Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*(8), 338-353.
265. Zadeh, L. A. (1994, March). Fuzzy logic, neural networks, and soft computing. *Communications of the ACM*, 37(3), pp. 77-84.
266. Zahariev, M. (2023, February 23). Law will always lag behind and gasp in trying to regulate new technologies (in Bulgarian). (G. Marinova, Interviewer) Блумбърг България.
267. Zhang, D., Maslej, N., Brynjolfsson, E., Etchemendy, J., Lyons, T., Manyika, J., . . . Perrault, R. (2022, March). *The AI index 2022 annual report*. Retrieved January 24, 2023, from https://aiindex.stanford.edu/https://aiindex.stanford.edu/wp-content/uploads/2022/03/2022-AI-Index-Report_Master.pdf
268. Zhang, L. (2020). *Initiatives in AI governance: Shaping the agenda for a responsible future December 2020*. Retrieved February 25, 2023, from <https://static1.squarespace.com/https://static1.squarespace.com/static/5ef0b24bc96ec4739e7275d3/t/5fb58df18fbd7f2b94b5b5cd/1605733874729/SRI+1+-+Initiatives+in+AI+Governance.pdf>

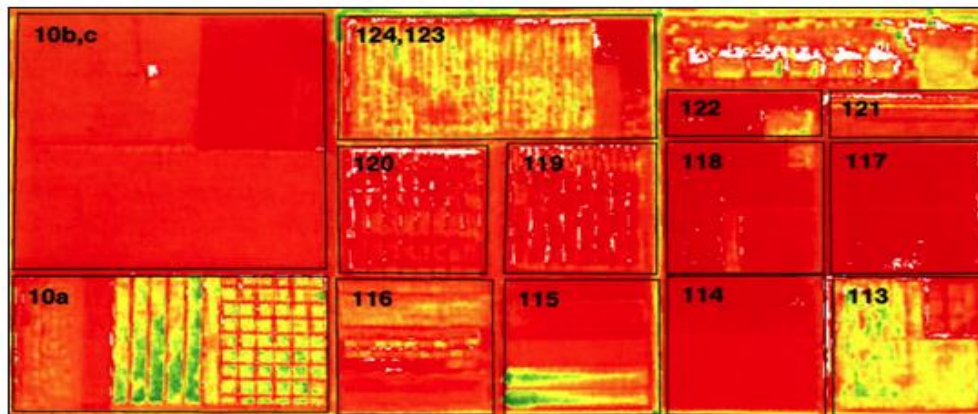
Appendixes

Appendix 1. AI for Precision Agriculture

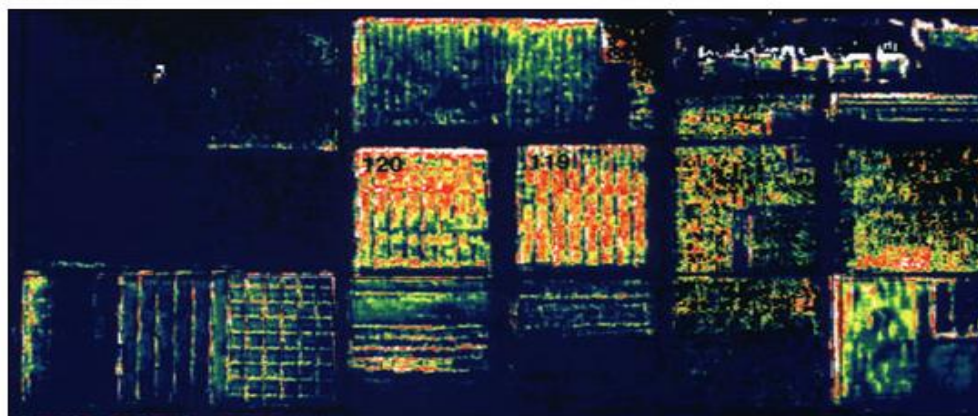
Precision agriculture is a management concept related to the observation, measurement, and search of the right place to grow crops. The figure below shows images acquired by the Daedalus sensor aboard a NASA aircraft (<https://earthobservatory.nasa.gov/images/1139/precision-farming>). The top image shows the crop density (dark blues and greens colours indicate lush vegetation, and red colour show areas without vegetation). The middle image is a map of measured water deficit - green and blue colours indicate wet soil and red areas are dry soil. The bottom image shows where crops are under serious stress (indicated by red and yellow pixels) and must be irrigated the following day.



Vegetation Density



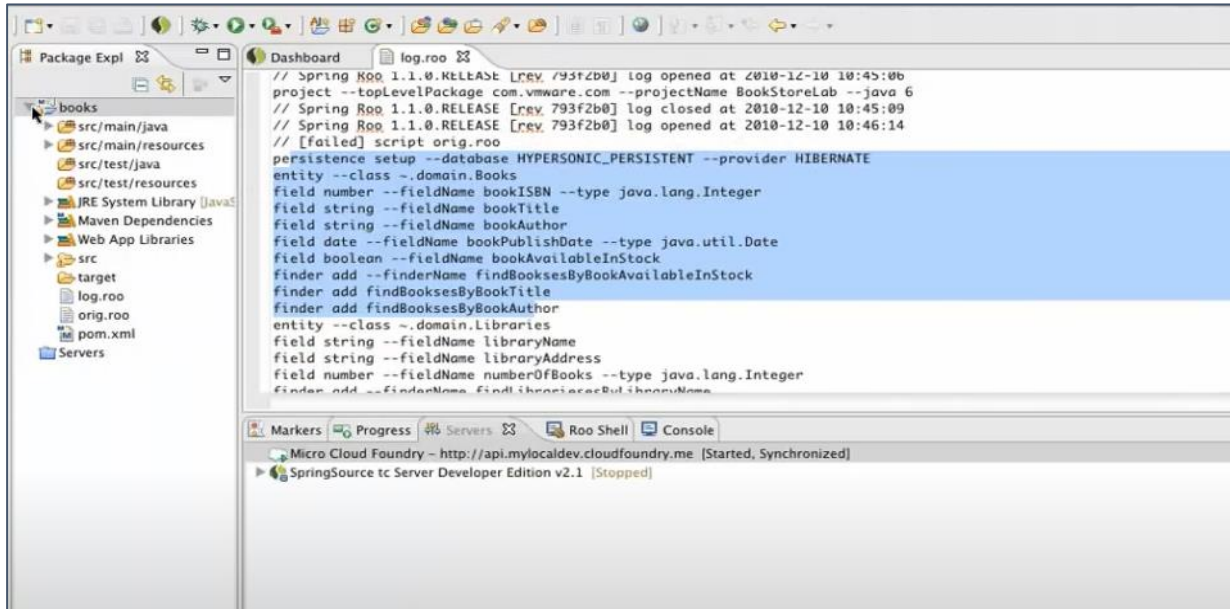
Water Deficit



Crop Stress

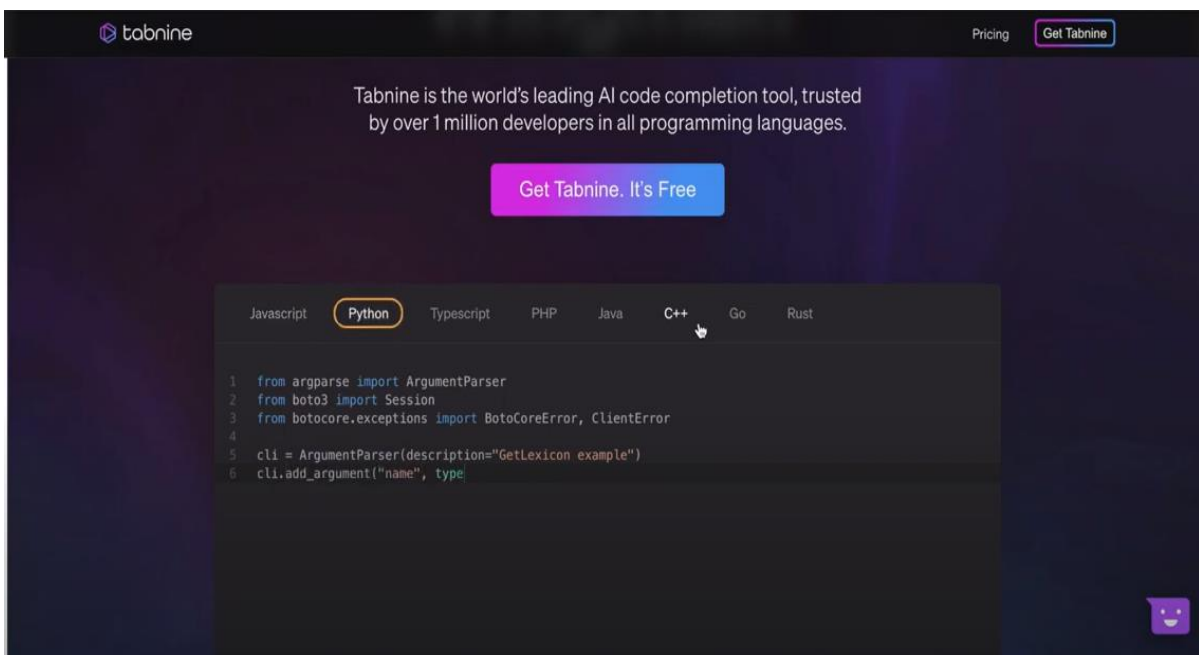
Appendix 2. AI for Industry Optimization

Application platforms as IBM Cloud (<https://www.ibm.com/cloud>), Predix (<https://www.ge.com/digital/iiot-platform>), or Cloud Foundry (<https://www.cloudfoundry.org/>) are designed to make organizations' s automation processes more effective and secure. The Cloud Foundry solution (on the image below) is an open-source project available on GitHub.



Appendix 3. AI Assistance for Program Code Synthesizing

Tabnine (<https://www.tabnine.com/>) is an AI assistant for software developers that learns from all the programming code in the organization and all other sources without code, such as Confluence, Wiki, code-related documentation, and then automates all parts of the workflow.



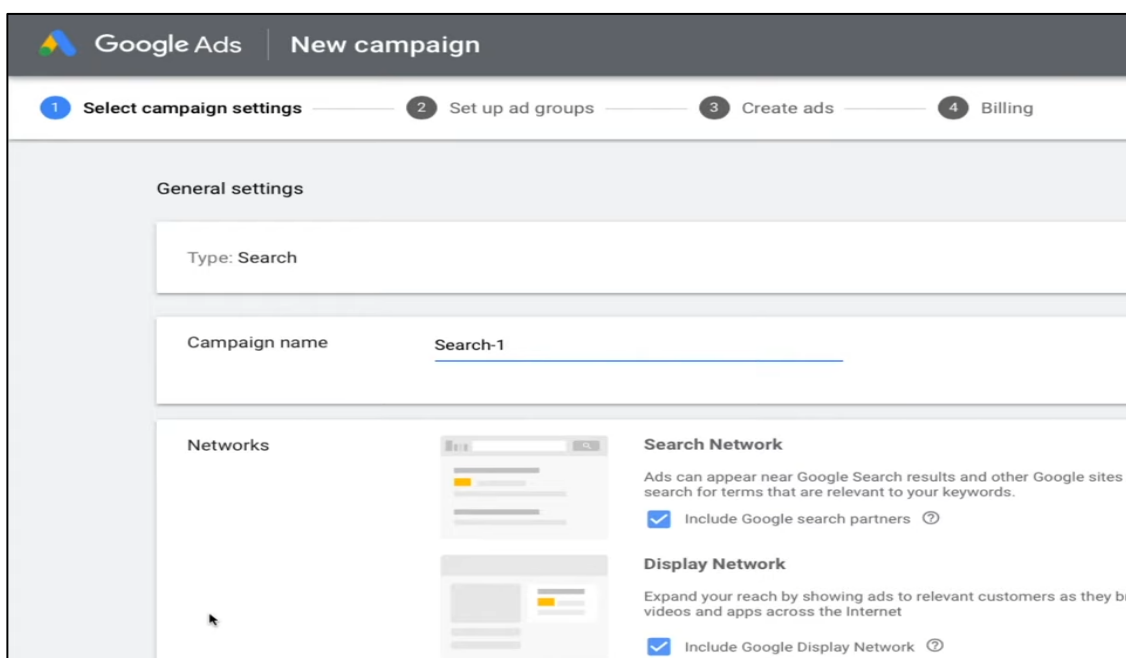
Appendix 4. Self-driving Assistance by AI Systems

Level 3 Mercedes-Benz's Drive Pilot system enables conditionally automated driving and frees time to drivers to browse the internet or to communicate, write messages and emails via in-car office tools. This Mercedes-Benz's ADAS system is now enhanced with the Intelligent Park Pilot (SAE Level 4) which allows automated self-parking.



Appendix 5. AI for Targeted Web Advertisements

Through the Google's AdSense program (<https://adsense.google.com/start/>) website publishers can create an revenue-generating advertising campaigns without a predefined goal targeted to the specific site content and audience.



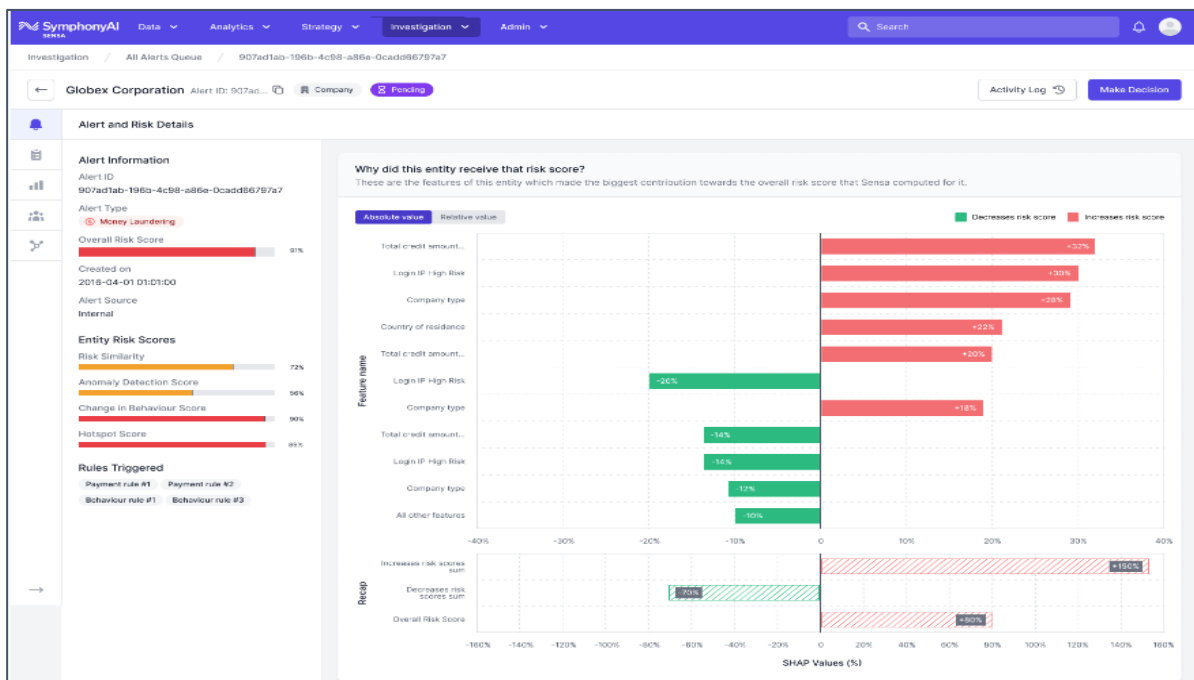
Appendix 6. Hospitality Serving Robots

The world's first hotel with robotic staff is the Japanese Henn naHotel (<https://group.hennahotel.com/>) where some of the robots are dinosaurs and holograms.



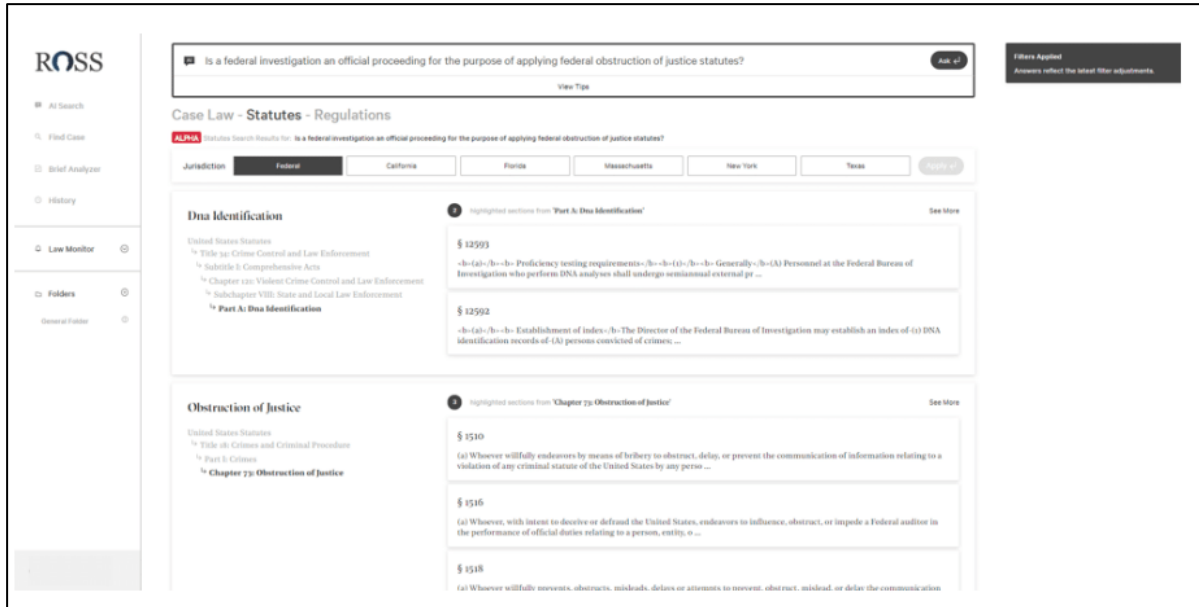
Appendix 7. AI for Anti-Money Laundering Detection

Symphony AyasdiAI Sensa (<https://symphonysensa.com/request-a-demo/>) is an AI application to discover hidden money laundering and financial crime activity. Financial institutions use this machine learning solution to analyse existing transaction data and identify anomalous behaviours in its correspondent banking operations.



Appendix 8. AI for Law Information Discovery

ROSS Intelligence Research Platform (<https://www.rossintelligence.com/features>) is trained on legal documents to use embedded words to recognize and understand the context, syntax and meaning of case law. It includes a full suite of American case law from all practices areas as well as statutes and regulations.



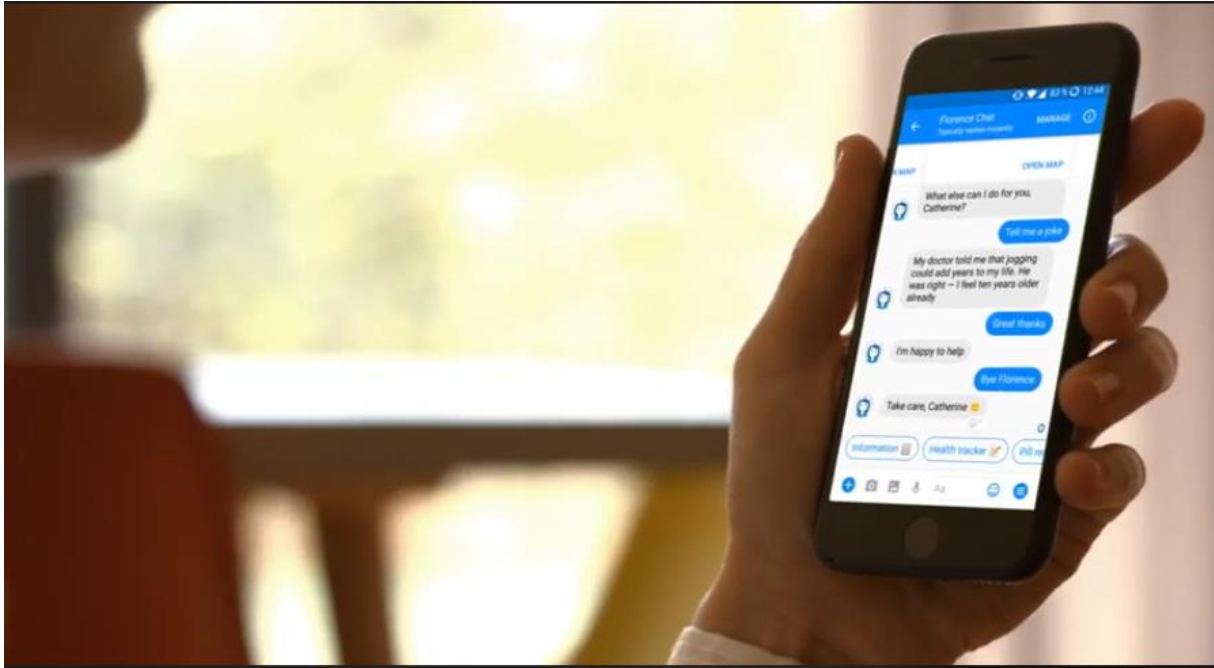
Appendix 9. Educational Robots

The Tico social robot developed by Adele Robots (<http://www.eu-robotics-sme.org/adele-robots-s-l/>) can interact with humans in different environments and can be used in education as a helper for teachers.



Appendix 10. Medical Chatbots

Florence (<https://florence.chat/>) is a healthcare chatbot which can remind users about their medication, check their symptoms, send them daily health tips, find a doctor for them or explain more about a disease.



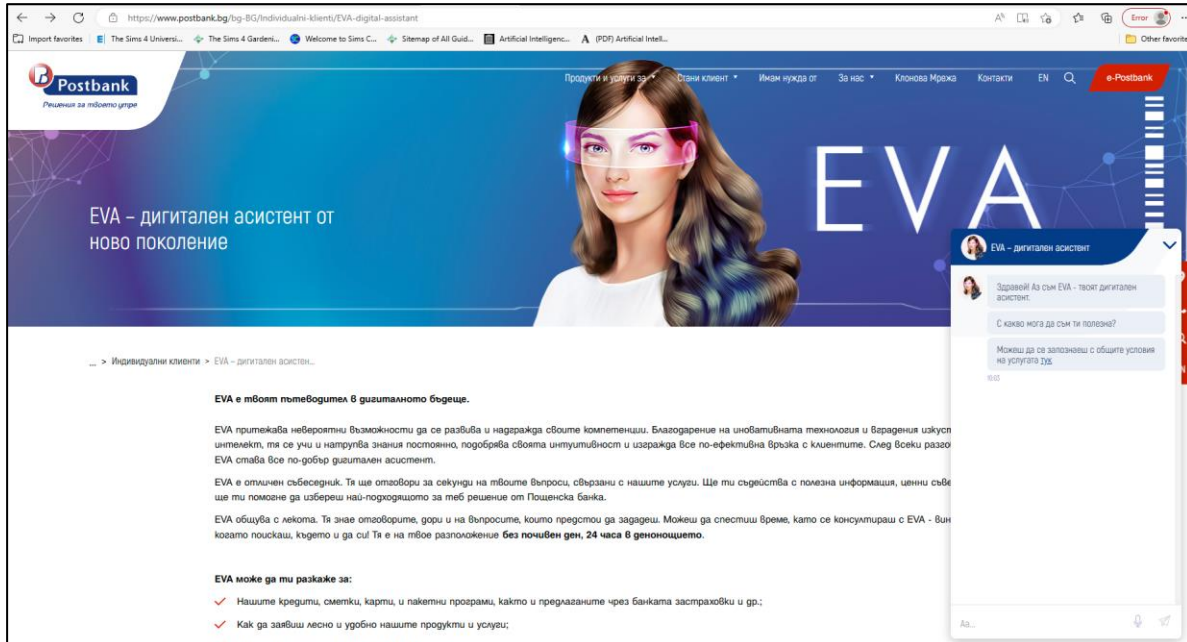
Appendix 11. Recruiting Robots

Furhat Robotics' Tengai social robot (<https://tengai.io/resources/tengai-robot/>) is an effective recruitment tool which can establish a connection with candidates during the job interview. In comparison to a chatbot, Tengai's social behaviour makes behavioural realism possible.



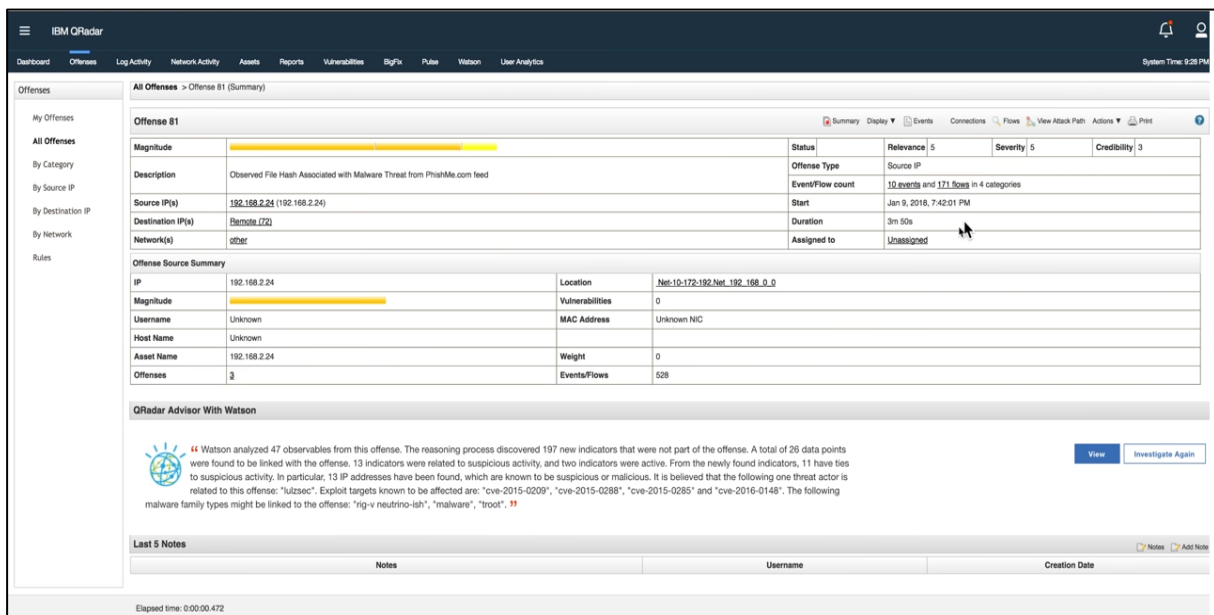
Appendix 12. Virtual Chatbot Assistants

EVA (<https://www.postbank.bg/bg-BG/Individualni-klienti/EVA-digital-assistant>) is a digital assistant, developed by Postbank, which assists customers with information and advice on choosing the right banking product.



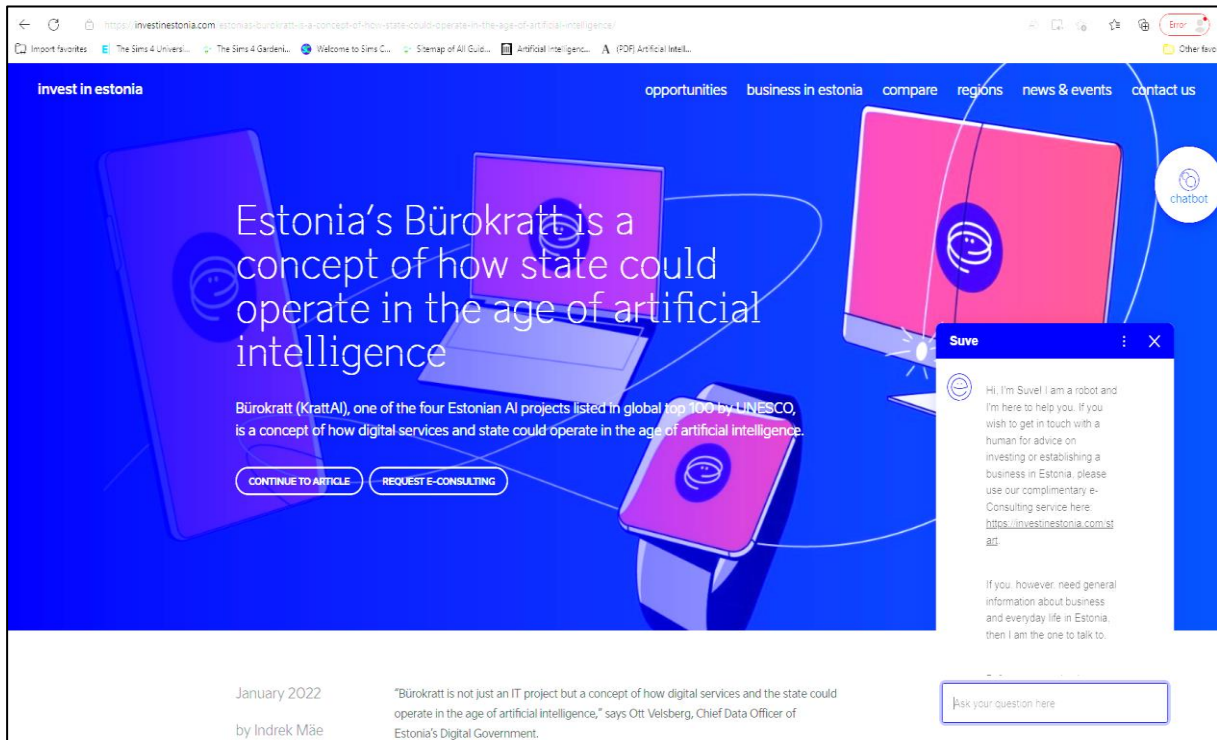
Appendix 13. AI Cybersecurity Solutions

IBM Security QRadar Advisor (<https://www.ibm.com/products/qradar-siem/addons#3071036>) can assess and reduce cyber risk incidents. Cognitive computing possibilities of IBM's Watson use advanced forms of artificial intelligence, including machine-learning algorithms and deep-learning networks, that get stronger and smarter over time.



Appendix 14. Virtual Interaction with Government Services

The Suve chatbot (<https://investinestonia.com/estonias-burokratt-is-a-concept-of-how-state-could-operate-in-the-age-of-artificial-intelligence/>) is a result of National Strategy for artificial intelligence of the Estonia's government. This digital service offers help in getting in touch with a human for advice on investing or establishing a business in Estonia.



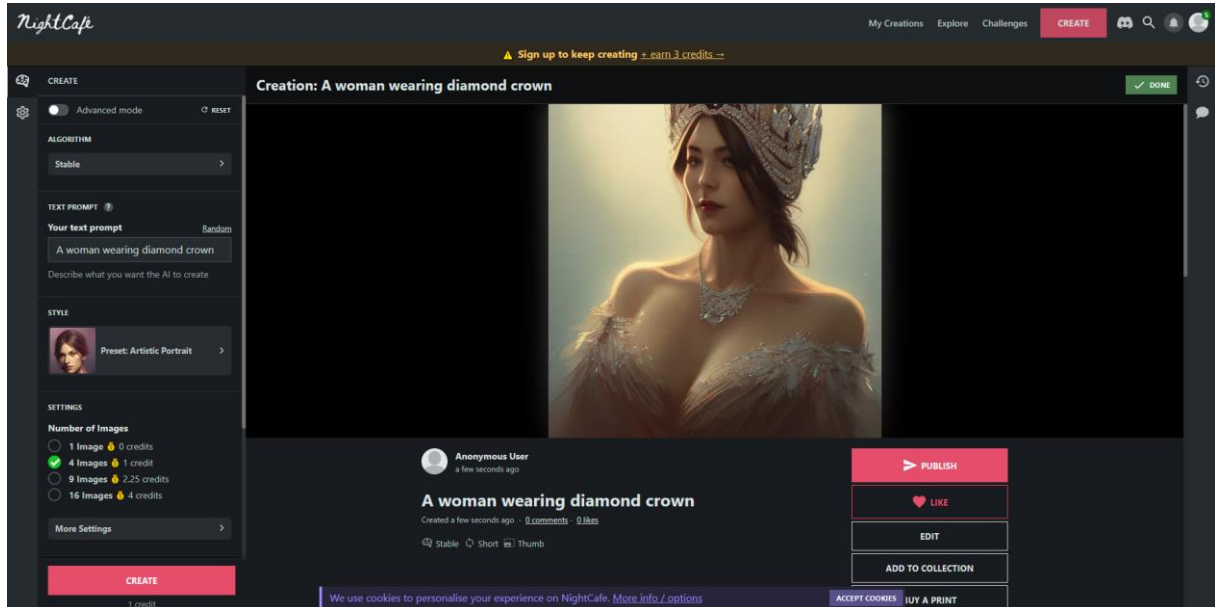
Appendix 15. Text Generating AI

The large language models based on transformer architecture consisting of billions of parameters trained on billions of words of text, can be used for grammar correction, creative writing, and generating realistic text. In this example, the ChatGPT-3 chatbot (<https://chat.openai.com/>) answers to a user's question about the time of the first human landing on the moon.



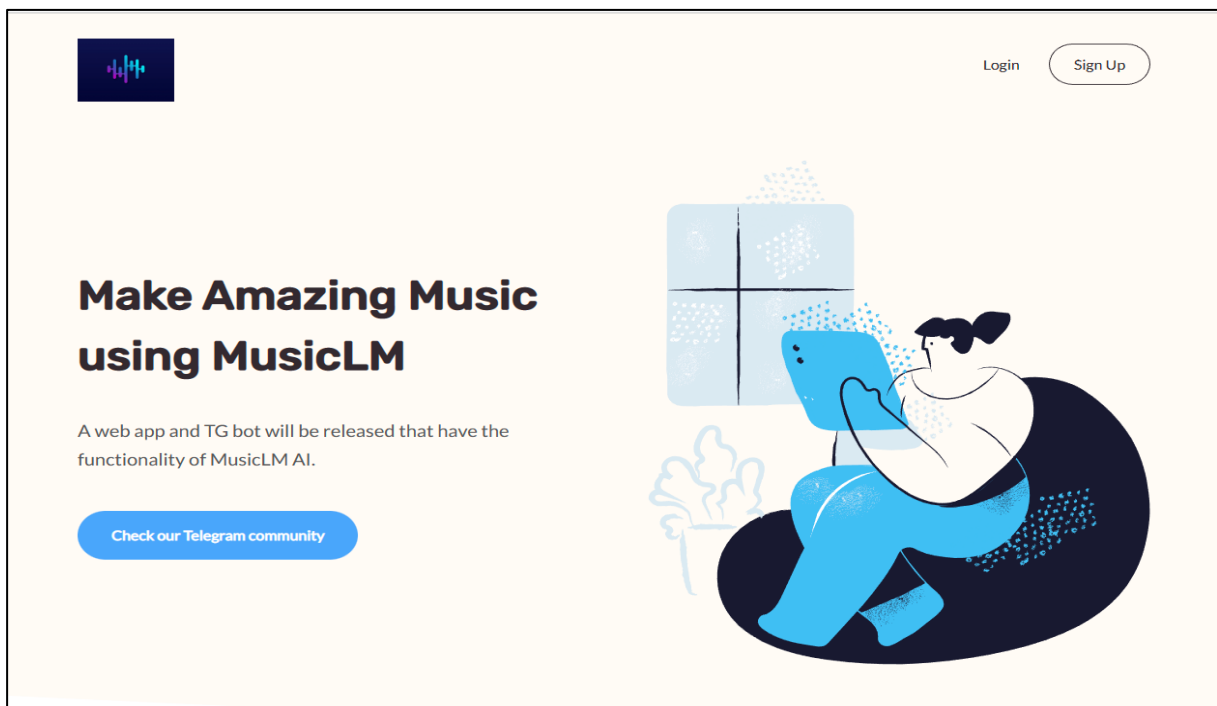
Appendix 16. Face Generating AI

The NightCafe's AI face generator (<https://creator.nightcafe.studio/ai-face-generator>) generates faces from text prompts and/or existing images. The example below shows artistic portrait on a woman wearing diamond crown.



Appendix 17. Music Generating AI

Google's MusicLM (<https://musiclm-ai.com/>) is a generative AI model that can create 24 KHz musical audio from text descriptions, such as "a calming harp melody backed by a distorted guitar riff." It can also transform a hummed melody into a different musical style and output music for several minutes.



Appendix 18. Image Generating AI

The GAN technology for generating images can be used to generate different kinds of images. The example below was produced by AI Image Generator (<https://www.aiimagegenerator.org/>) given the prompt “a stained glass window with an image of a pink strawberry.”

