ARTIFICIAL INTELLIGENCE FROM A TECHNICAL PERSPECTIVE

Sokolov Artem
Professor, Doctor of Technical Sciences, Department of Cybersecurity, National University "Odesa Law Academy"
ORCID ID: 0000-0003-0283-7229

Abstract. This monograph section provides a comprehensive analysis of the evolution, technological foundations, and current applications of artificial intelligence (AI), with a particular focus on its role in cybersecurity. We present a historical overview of AI development, tracing its path from the early conceptual ideas of the mid-20th century to the emergence of modern deep learning technologies, generative models, and large-scale Transformer architectures. Special attention is given to the critical technological breakthroughs that enabled the rapid growth of AI capabilities, including advances in computing hardware, neural network architectures, and algorithmic training methods.

We examine the technical foundations of AI systems, focusing on the architecture and operation of artificial neurons and neural networks. The discussion covers core machine learning and deep learning techniques, with particular attention to natural language processing models such as Transformers, BERT (Bidirectional Encoder Representations from Transformers), and GPT (Generative Pre-trained Transformer). The role of generative adversarial networks in advancing creative and synthetic AI applications is also analyzed, with a focus on their technical mechanisms and real-world uses.

The concept of explainable AI is considered, addressing the growing need for transparency, interpretability, and accountability in the deployment of complex AI systems. Various technical approaches to model explainability are discussed, including their strengths, limitations, and significance for trust-building in critical domains.

The integration of artificial intelligence into cybersecurity is presented as a transformative force, significantly enhancing capabilities in threat detection, anomaly analysis, intelligent event processing, cryptography, steganography, and the development of autonomous defense agents. Through the lens of cybersecurity, we underscore AI's pivotal role as a foundation for proactive, resilient, and adaptive digital protection strategies in an increasingly interconnected and volatile technological environment.

Keywords: Artificial Intelligence, Neural Networks, Machine Learning, Deep Learning, Transformers, Generative Adversarial Networks (GANs), Explainable AI (XAI), Cybersecurity, Threat Detection, Cryptography, Steganography, Anomaly Detection, Autonomous Cyber Defense.

Research methods. The research presented is based on a combination of theoretical and analytical scientific methods, selected according to the interdisciplinary nature of the subject. The historical development of artificial intelligence was explored using methods of analysis and synthesis, allowing for the identification of key milestones and technological trends that shaped the evolution of AI systems.

A comparative analysis method was employed to examine and contrast different technical architectures of AI, including neural networks, Transformer models, generative adversarial networks (GANs), and explainable artificial intelligence (XAI) approaches. Inductive and

deductive reasoning were used to generalize findings from specific case studies and technological examples, enabling the formulation of broader conclusions about the role of AI in digital transformation.

Modeling techniques were applied to describe the internal mechanisms of AI functioning, such as the operation of artificial neurons, learning algorithms, attention mechanisms, and the structure of autonomous cyber-defense agents. The research adopted a systems approach to assess AI as an integral part of larger digital ecosystems, particularly in the context of cybersecurity infrastructures.

Finally, expert evaluation and theoretical forecasting methods were used to analyze current challenges and predict future developments in the application of AI technologies to threat detection, cryptography, steganography, and anomaly analysis. This comprehensive methodological framework ensured a thorough and structured exploration of both the technical and applied aspects of artificial intelligence.

Introduction: The place of artificial intelligence in the digital transformation of society. Artificial intelligence (AI) has long been one of the most important technologies that determine the development of the modern world. Its evolution, spanning several decades, reflects achievements in science and technology and profound changes in social, economic, and legal structures [1]. The beginning of this story dates back to the middle of the 20th century, when scientists first began to think about creating machines capable of performing intellectual tasks that were previously inherent only to humans. Since then, AI has come a long way, from the first experiments and theoretical models to modern deep learning algorithms and neural networks, which today penetrate all areas of life, from medicine to finance and justice. However, with the development of AI, new challenges also arise, particularly in matters of ethics, privacy, and human rights, which require careful analysis and the development of appropriate legal norms.

The history of the development of artificial intelligence not only opens up technological horizons for us, but also raises important questions about how humanity should interact with this new reality.

Depending on the type of tasks that AI performs, it can be divided into different categories.

However, it is worth noting that AI is still in the early stages of development. In the future, it will look and behave very differently from what it does today.

Types of AI depend on the level of intelligence they demonstrate. Three main categories can be distinguished (Fig. 1) [2].

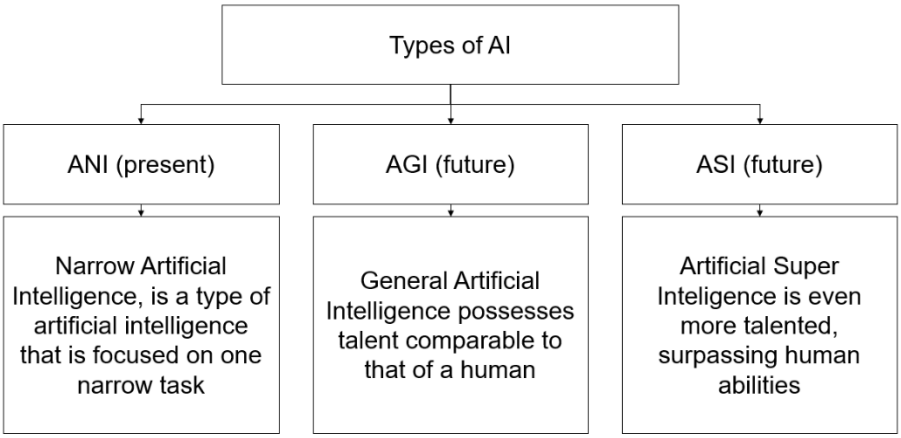

Fig. 1. *AI Categories*

Artificial Narrow Intelligence (ANI), also known as weak AI, specializes in performing specific tasks and has a limited range of abilities. This is the only type of AI we deal with today. Narrow AI is the basis of many everyday technologies, such as Google Assistant, Google Translate, Siri, Cortana, and Alexa. All of these systems use Natural Language Processing (NLP), which allows them to interact with humans naturally and intuitively. AI also finds wide application in medicine, where it is used to diagnose diseases such as cancer with exceptional accuracy, reproducing human perception and thinking.

Artificial General Intelligence (AGI) is a type of AI that has human-like capabilities. However, AGI is still in the development stage. Since the human brain is the model for creating such intelligence, achieving AGI will take more time. The lack of complete knowledge of how the brain functions makes it difficult to develop this type of AI. However, history shows that humanity is prone to creating technologies that can become existentially dangerous. When we achieve AGI, humans must be prepared for the consequences that this may bring.

Artificial Super Intelligence (ASI) is a future prospect that we are only just getting closer to. To reach this level, AI must surpass human ability in all areas. Superintelligence will be able to successfully solve tasks that are currently considered purely human, such as art, decision-making, and emotional relationships. This raises the question: have humans truly mastered the art of emotional relationships and decision-making? If not, is it possible that AI with superintelligence will eventually be able to effectively manage areas where humanity still struggles?

Today, AI has become one of the main driving forces of global digitalization, as it allows for a significant increase in the efficiency and speed of data processing, as well as optimizing numerous processes in various industries. Digitalization, which covers all spheres of life, from the economy to education and healthcare, requires the integration of the latest technologies, and AI is a key tool to achieve this goal.

AI not only stimulates the development of digital infrastructures but also transforms how people and technology interact. Thanks to its capabilities for automation and machine learning, AI allows for a significant increase in the efficiency of data processing and optimizes processes in various industries. From economics to medicine, from education to manufacturing, AI opens up new opportunities for creating innovative solutions and radically changes approaches to old problems. This transition to a digital society is an important step in the development of humanity,

pushing civilization to a new stage of progress, similar to previous revolutionary changes in production processes, social structures, and culture.

In the financial sector [3], for example, AI not only automates transactions but also improves risk management and fraud prevention.

In medicine [4], it allows for the diagnosis of complex diseases with previously unimaginable accuracy and the creation of personalized treatments. These are just a few examples of how AI is becoming a critical element not only for global digitalization but also for the evolution of human institutions.

AI is also an important factor in the development of education and science [5]. Personalized learning, the automation of scientific research, and the processing of big data are opening up new horizons for the intellectual development of humanity. However, along with the capabilities of AI, new challenges arise, such as ethical, social, and legal, which require a comprehensive approach to regulation and control. After all, as with previous stages of the technological revolution, the changes brought by AI do not come without risks to privacy, security, and social stability.

Thus, AI is not only a driving force for digital transformation, but also reflects deeper changes in the development of human civilization. It is a stage at which humanity is faced with new opportunities and the responsibility for how these opportunities will be used in the future.

Information technology (IT) has become an integral part of our daily lives, significantly changing all areas of activity, from business and education to medicine, culture, and law enforcement. The development of digital technologies has brought with it numerous benefits, but at the same time has posed new challenges to society, among which the issue of cybersecurity occupies a special place. The integration of IT into all aspects of life has contributed not only to the creation of convenient and effective solutions but also to a significant increase in the number of cyber threats arising from the use of big data, artificial intelligence, and other new technologies.

Every year, cyber threats become more complex and dangerous as the number of devices connected to networks, the volume of data transmitted, and the degree of automation of processes increase. This increases the vulnerability of systems that process confidential and personal information, and also creates new opportunities for cybercriminals. Systems that interact with AI are especially vulnerable. After all, advanced artificial intelligence algorithms can be used both for protection and for attacks. For example, AI can optimize "phishing" attacks, creating increasingly convincing fake messages and sites, and automate the attack process, which makes it more difficult to detect cybercriminals. In addition, with the increasing use of AI in organizations, the question arises about the security of artificial intelligence systems themselves [6…9]. Machine learning algorithms, like other automated systems, can be vulnerable to manipulation or attacks aimed at distorting their learning. If such systems are compromised, cybercriminals can alter the results of their work, which will have serious consequences for data security and confidentiality.

In addition to technological challenges, AI also poses new challenges for the legal system [10]. Since artificial intelligence can make automated decisions, the question of liability for these decisions arises. For example, who is liable if an AI system makes a decision that causes harm to an individual or organization? This question is important for judicial practice, since laws governing liability cannot always be adapted to new realities where decisions are made by algorithms, not people. Also, since AI is able to analyze huge amounts of data, there is a risk of violating the rights to confidentiality and privacy. Creating legislative mechanisms that ensure the protection of personal information in the context of the comprehensive use of AI requires new approaches to

legal regulation. This includes creating effective rules to ensure data security and processing, as well as protecting human rights in the context of the creation and development of AI.

Part 1. A brief history of the development of artificial intelligence

To better understand the current challenges and opportunities that arise in connection with the development of AI, it is important to familiarize yourself with its history. The development of AI is not just a technological progress, but a whole path that includes numerous stages, from the first concepts to modern achievements. The history of AI allows us to understand how changes in science and technology have affected society and what philosophical, ethical, and legal issues have become relevant at different stages of this development. Studying the past of artificial intelligence helps us understand what lessons we can learn from previous attempts to create machines capable of intelligent activity, and how this knowledge shapes our ideas about the future of technology.

To understand the origins of the concept of artificial intelligence, it is necessary to consider the scientific and technological advances that took place in the mid-twentieth century. One of the defining factors that contributed to the emergence of AI was the revolution in computing technologies, in particular Moore's Law [11]. In 1965, Gordon Moore, one of the founders of Intel, formulated his famous law, according to which the number of transistors on a microcircuit doubles every two years, which leads to a constant increase in the computing power of computers while reducing the cost of these devices. This law became the basis of technological progress in computing and accelerated the development of scientific research in the field of artificial intelligence. Equally important for the development of computing is Bell's Law, formulated in the 1970s by Chester Bell [12], which states that every decade a new class of computing technology appears on the market, changing the paradigm of data processing. Going beyond the period under consideration, for the sake of completeness, we note that in the 1980s and 1990s, bulky computers were replaced by desktop computers, which became accessible to the mass user and provided new opportunities for the development of software products, including AI algorithms. In the 2000s, laptops and mobile devices, thanks to increased mobility and computing power, opened up new horizons for personal and business applications of computer technologies.

In the early 2010s, the Internet of Things (IoT) [13] and robotics became new revolutionary trends, allowing for the collection of vast amounts of data from the real world, which became the basis for the development of more complex AI algorithms. But in our 20s, we witnessed the emergence of an AI technology that is already changing the world. This technology is gaining increasing importance, as its implementation in various fields (from medicine to education, from manufacturing to cybersecurity) opens up new development opportunities, while at the same time posing new challenges to society, both ethical and legal.

Thus, technological changes in computers, from vacuum tubes to microprocessors, created the foundation for the further development of artificial intelligence. Technological advances allowed the development of new models for creating intelligent systems that could imitate human cognitive processes, such as speech and image recognition, decision-making based on large amounts of data, etc.

In the context of this development, we cannot ignore Alan Turing [14] (Fig. 2), one of the founders of the theory of artificial intelligence, whose work became the key to understanding the very concept of “machine thinking”.

Fig. 2. *Portrait of Alan Turing created by generative AI*

Turing is best known for his Turing machine, an abstract mathematical model of the process of computation that would later define the foundations of computer science. In 1950, he published his famous paper, "Computing Machines and Intelligence" [15] in which he proposed the Turing Test as a criterion for determining whether a machine could be considered intelligent. However, Turing faced great difficulty in understanding how to organize computations so that the machine could act "intelligently" and how to effectively solve complex problems, such as pattern recognition or natural language processing. The idea was to develop algorithms that would have the ability to adapt, but at that time they were difficult to implement due to limitations in computing power and the lack of big data. Turing also proposed a solution to the problem of emulating human thinking using algorithms, but his concepts were often criticized for being impractical and beyond the scope of existing technology.

The challenges that scientists faced during this period included not only the limitations of technical capabilities, but also fundamental questions about cognition and intelligence. How can a machine be given the ability to understand, interpret, and adapt knowledge to new conditions? What algorithms can model such flexibility? Although the solution has not yet been found, the questions themselves have become a catalyst for further research, leading to revolutionary changes in the future. The period from 1950 to 1980 marked the beginning of a transformation of ideas [16] that today form the basis of the development of modern artificial intelligence. The first conceptual and theoretical works laid the foundations for the further development of technologies that would allow us to achieve what once seemed impossible.

It should be noted that the 1980s were an important stage in the development of artificial intelligence. This period was marked by intensive work on expert systems, which aimed to imitate the decision-making process of a human expert in a certain field. One of the most significant expert systems of that time was MYCIN [17], developed in the 1970s at Stanford University. MYCIN was able to diagnose infectious diseases and suggest treatments based on a knowledge base that was entered by experts. However, these systems had significant limitations: their functionality was limited by rigidly defined rules, and they could not adapt to new or unknown situations. In the 1980s, one of the important figures in the development of AI, John McCarthy, proposed the concept of “Circumscription”, which became the basis for building programs capable of solving complex logical problems [18]. McCarthy, one of the founders of the AI research group at Stanford

University, is particularly known for his work on the LISP programming language [19], which remained the primary language for AI for many years.

Despite the high hopes for expert systems, researchers faced serious problems, such as the difficulty of accumulating large amounts of data to properly train the systems and the excessive complexity of maintaining the rules. However, this period also became important for the search for alternative approaches to solving AI problems. Among the greatest achievements of this period was the emergence of neural networks, which became the basis for further achievements in the field of AI. In the 1980s, engineer Jeffrey Hinton, together with his colleagues, initiated work on backpropagation of errors [20], which allowed neural networks to learn efficiently on large data sets. However, the development of neural networks stalled for several decades due to limited computing power and the lack of large data sets for training models.

In the mid-1990s, the concept of machine learning was also developed using statistical methods such as support vector machines (SVM) [20], which helped to solve a number of practical problems in the field of data and pattern processing. However, the problem remained the same: while these methods were effective in certain cases, they could not achieve the desired results in more complex situations, such as natural language processing or image recognition.

Although the concept of deep learning was known even earlier, the real breakthrough in its application occurred in the mid-2000s. It was then, thanks to the work of Yann LeCun, that deep neural networks, in particular CNN (Convolutional Neural Network) [21], began to show impressive results.

LeCun's research on deep learning methods became the basis for significant advances in the field of computer vision. However, despite the successes, at that time, such systems were still limited by technological capabilities and could not work on large amounts of data.

Since the 2010s, deep learning technologies have been developing rapidly, largely due to the significant increase in computing power and access to large data sets. Deep learning systems have been able to effectively perform complex tasks such as automatic translation, natural language processing, face recognition, and many others. In particular, in 2012, the achievement of Alexey Krizhevsky and his colleagues, who developed the AlexNet model [22] for the task of image classification, was a breakthrough in the field of computer vision. AlexNet was the first model to achieve outstanding results in the international ImageNet competition, marking a pivotal moment in the evolution of deep learning.

As a result, the period 1980-2010 was a stage when artificial intelligence went from theoretical expert systems to real-world applications of deep learning capable of solving complex problems. This period was marked by the search for new methods, the spread of ideas, and the achievement of scientific breakthroughs that became the basis for the further development of AI in the following decades.

The 2020s have witnessed a massive shift in the landscape of artificial intelligence technologies. The emergence of new innovative approaches, such as generative models and deep learning, has led to breakthroughs that are actively transforming various industries, from medicine and finance to the creative industries and cybersecurity [23]. Today, AI technologies are actively implemented in everyday life, changing the way people, businesses, and governments work. At the same time, this gives rise to a number of global challenges that question not only ethical and legal norms, but also the basic principles of the functioning of society. One of the most significant achievements of the last decade has been generative models (such as Generative Adversarial Networks or GANs, and diffusion models) [24]. These models are capable of not only analyzing

large amounts of data but also creating new, previously non-existent content samples. For example, generative model algorithms can generate images, music, text, and even videos that look so realistic that they can be mistaken for human creation. This has given rise to new forms of art, entertainment, and has also changed approaches to marketing and advertising, where artificially created images have become an integral part of visual content.

However, this ability of AI to create "new realities" carries with it risks associated with the possibility of manipulating information. The ability to generate fake videos or texts (for example, deepfakes) raises questions for society about the reliability of information, as well as protection against manipulation, especially in the political and media spheres.

Another major achievement in recent years has been the improvement of deep learning algorithms [25], which has become possible due to increased computing power and the availability of big data. Deep learning models, in particular neural networks based on multilayer architectures, are capable of achieving high results in image recognition, natural language processing, and even strategic planning.

Deep learning has enabled more accurate medical diagnosis, robotics, and automated systems like autonomous vehicles. However, it also poses new challenges, from the "black box" problem (when an algorithm decides without a clear explanation of how it works) to the need to protect deep learning models from attacks such as adversarial attacks that can manipulate results in real time.

One of the most significant changes in everyday life has been the widespread adoption of AI in the form of personal assistants such as Google Assistant, Siri, and Alexa, which are actively used to control household devices, communicate with users, and provide information. Such technologies have significantly changed the way people interact with technology and allow them to automate a number of tasks that previously required human intervention. From automating everyday processes to developing robots that can perform complex physical and cognitive tasks, AI is making our world more integrated and efficient.

However, with the widespread adoption of such technologies, serious privacy and security issues arise. Systems that use AI for monitoring, data collection, and decision-making can be subject to abuse and interference, and raise questions about who owns the data and how it is used.

As artificial intelligence becomes more pervasive, new global challenges emerge, not only for technology but also for society as a whole. One of the most important challenges is the ethics of AI. How can we ensure that technology does not violate human rights or become a tool for manipulation or discrimination [26]? How can we ensure that algorithms are transparent and accountable when their decisions can have a huge impact on people's lives?

AI also threatens traditional forms of employment. From automating manufacturing processes to replacing professionals in fields such as law, medicine, and finance, machines can perform many tasks previously only available to humans. This requires adapting curricula and retraining workers for new conditions.

An equally important challenge is the regulation of AI technologies, which is becoming extremely important for ensuring the safety and ethics of their use. At the global level, issues of ethics, human rights, and privacy are already being discussed by governments, but there is still no solution for universal standards for the use of AI.

Artificial intelligence in the 2020s has become a technology that is not only developing but also actively integrated into all aspects of life in modern civilization. From generative models to serious challenges in the field of ethics and security, this period has become a turning point, when

AI is not just changing science and technology, but also posing new questions for humanity that have no simple answers. Understanding these changes and being ready to adapt to them is important for artificial intelligence to continue to benefit society, rather than causing unpredictable consequences.

Part 2. Technical foundations of AI functioning

Neural networks are a class of machine learning algorithms inspired by the human brain and are the foundation for the development of artificial intelligence. They consist of numerous neurons that are interconnected and perform operations on input data. The most important components of a neural network are its architecture, a training algorithm, and activation functions that determine how the neurons will respond to input information.

An artificial neuron [27] is the basic element of a neural network, inspired by a biological neuron in the brain. With their ability to adapt to different types of input data and process them nonlinearly, neurons provide efficient pattern recognition and perform tasks such as image classification, text processing, and more. Each neuron is responsible for its part of the data processing, and together they form powerful networks that can solve complex problems with a high level of accuracy.

According to the principles on which neural networks are built, an artificial neuron has three main functions: to receive input signals, process them, and pass the result to the next neuron or the output.

Suppose an artificial neuron is like a small computational "node" that receives several input signals, processes them, and outputs a result.

For example, we are trying to train a network to recognize photos of cats. In this case, the input data for the neurons will be pixels of an image of a cat.

1. Inputs: Each pixel in the image is an input to the neuron. If the pixel is light, the value will be high, if it is dark, the value will be low. For example, a single neuron might receive pixels at coordinates (i, j) in an image of a cat.

2. Weights and bias: Each input has a weight, a coefficient that determines its importance. For example, if a pixel in an image of a cat is bright white, it may be important for recognizing the shape of a cat, so the pixel weight will be higher.

3. Bias is an additional parameter that allows the neuron to adjust its response regardless of the input signals. Bias helps the model adapt better to different input variations.

4. Input processing: The neuron multiplies the input signals by their weights and adds a bias. The sum is then passed through an activation function. The activation function transforms the result into a specific output, determining whether the signal will be passed on or not. For example, the activation function could be a ReLU (Rectified Linear Unit), which gives a positive result if the sum of the signals is greater than zero, and zero if less.

5. Activation Function: The most common activation functions include:

– Sigmoid: Returns a value between 0 and 1, allowing the neuron to determine the probability that a cat is present in an image.

– ReLU: Returns 0 if the sum of the inputs is less than zero, or the sum itself if it is greater than zero. This is useful for handling nonlinear relationships in data.

– Tanh: Similar to Sigmoid, but returns a value between -1 and 1, allowing for better handling of symmetric data.

6. Output: When a neuron processes its input through an activation function, it passes the resulting value to subsequent neurons or to the output. For example, if a neuron is working in a network that is designed to determine whether a cat is present in an image, the output might be the probability that the image contains a cat.

Artificial neurons are combined into layers. Each layer of neurons interacts with the next layer, where data processing becomes increasingly complex and abstract. In simple networks, there may be one layer, and in deep neural networks, there may be dozens and hundreds of layers. Each neuron in the hidden layers of the network is responsible for learning certain features or patterns from the data, which allows the network to acquire more complex connections and make complex predictions. An example of a neural network is shown in Fig. 3.

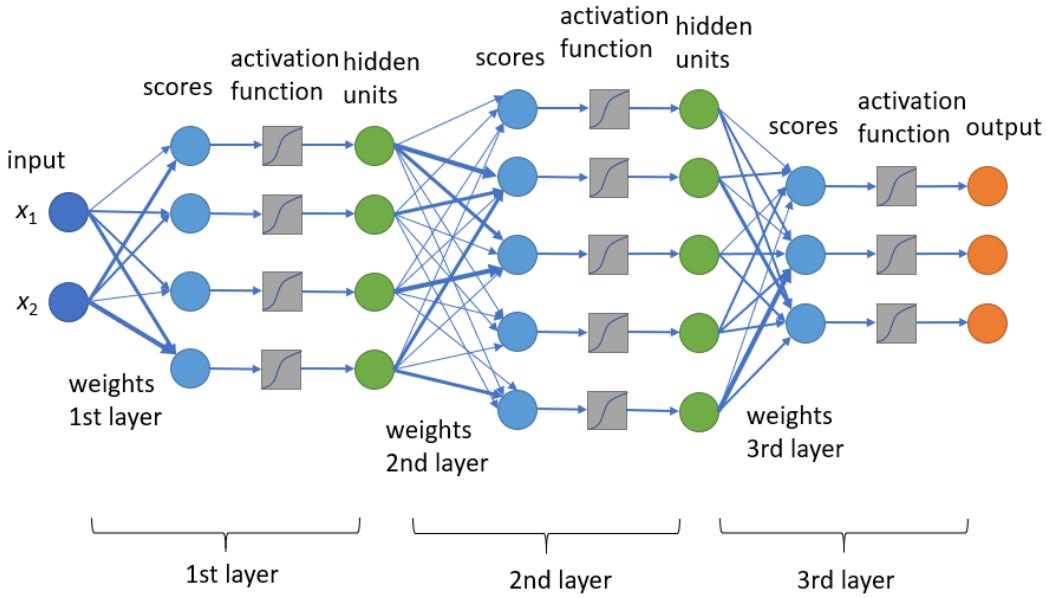

Fig. 3. *Example of a neural network*

Let us consider the main features of neural network technology:

1. Multilayer architecture: Modern neural networks typically have several layers of neurons. The first layer is the input layer, which receives data, then there are several hidden layers where processing occurs, and the final layer is the output layer, which generates the result. The number of layers and neurons in each layer determines the complexity and power of the network. Multilayer architectures (e.g., MLP, CNN, RNN) allow networks to solve complex problems, such as pattern recognition or natural language processing.

2. Training and adaptation: Training a neural network is performed using a backpropagation error algorithm, which allows adjusting the weights of connections between neurons based on the difference between the prediction and the actual result. Training can be supervised, where reference data exists, or unsupervised, where the network looks for structures in the data without external supervision.

3. Activation functions: In order for a neuron to be able to perform nonlinear information processing, activation functions are used. As it was mentioned before, the most common are ReLU

(Rectified Linear Unit), Sigmoid, and Tanh. The choice of activation function affects the network's ability to learn and its efficiency in processing data.

4. Processing large amounts of data: Neural networks are very effective at working with large amounts of data, as they can automatically find patterns and correlations without the need for manual feature extraction. This makes neural networks particularly powerful for processing images, text, and sound.

5. Generation and recognition: Due to the ability of neural networks to generate new data, they are actively used to create fake images, synthesize text, and also to recognize objects or detect anomalies in data.

Neural networks require specialized hardware [28] to efficiently process and train large models. Because training modern neural networks is computationally intensive, traditional central processing units (CPUs) often cannot handle the workload, so hardware platforms such as graphics processing units (GPUs) and tensor processing units (TPUs) are used.

1. Graphics processing units: GPUs have become the primary hardware for training neural networks because of their ability to handle parallel computations. Unlike CPUs, which perform serial computations, GPUs can perform thousands of computations simultaneously, making them ideal for tasks that require a large number of parallel operations, such as training deep neural networks. Popular GPU manufacturers for AI are NVIDIA and AMD. Tools such as CUDA (from NVIDIA) allow you to optimize code for GPUs, which significantly speeds up the process of training models.

2. Tensor Processing Units (TPUs): Developed by Google, Tensor Processing Units (TPUs) are even more specialized chips that are optimized for processing large amounts of matrix operations, which are the core operations in many neural networks. TPUs are used for large deep learning models, especially in tasks that require a large number of matrix operations, such as generative models or transformers. Using TPUs can reduce model training time and power consumption.

3. Cloud-based models: Because training complex neural networks can be very resource-intensive, many companies choose to use cloud computing platforms such as Google Cloud, AWS, or Microsoft Azure. These services provide access to powerful GPUs and TPUs through APIs, allowing developers to train neural networks without the need for physical servers.

4. Application-Specific Integrated Circuits (ASICs): ASIC hardware is used for specialized computing. They can be configured to perform only a specific set of operations, such as training neural networks, which allows for even greater efficiency and reduced power consumption. This approach is often used in supercomputers or large data centers.

5. Distributed computing: In cases where training models on a single machine is not possible due to resource constraints, distributed computing technology is used. Different processors or graphics cards can work together to perform model training, which allows the load to be distributed across multiple pieces of hardware and speeds up the process.

Let's consider the architecture of modern neural networks: types, training algorithms, and implementation examples.

Perceptron [29]. The perceptron (Fig. 4) was one of the first types of neural networks, proposed in 1958 by Frank Rosenblatt. It was a single-layer neural network used for binary classification, where each neuron accepted certain inputs and made decisions based on a threshold. The model was limited in its capabilities, as it could not solve more complex problems due to the restriction of having only one layer.

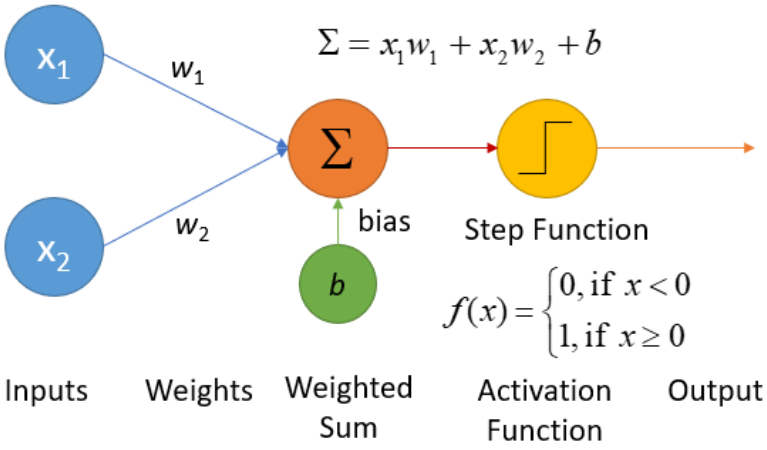

Fig. 4. *The perceptron*

However, the perceptron became the basis for the development of more complex neural networks. It served as a basic example for studying the concept of artificial neurons and continues to be used in certain simple problems. Today, the perceptron is an important tool for teaching the basic principles of neural networks.

Multilayer perceptron (MLP) [30]. The MLP (Fig. 5) is a development of the classical perceptron. It was designed to solve more complex problems, as it has several layers of neurons, between which there are connections. The backpropagation algorithm, which was proposed in 1986 by Jeffrey Hinton and his colleagues, has made it possible to train multilayer networks efficiently.

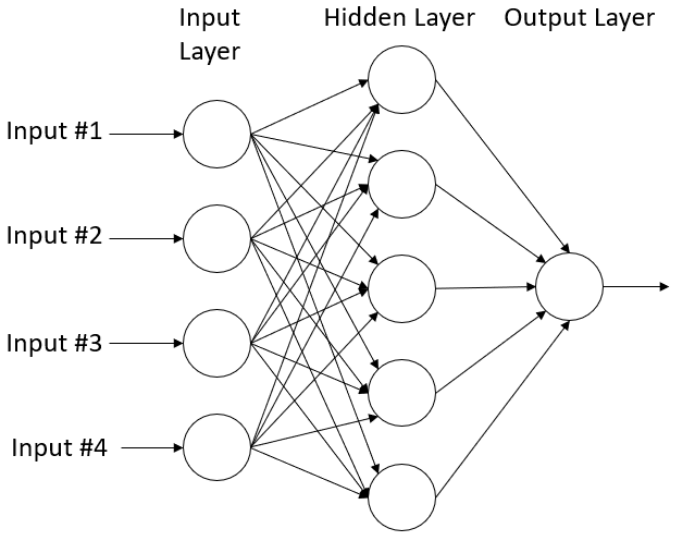

Fig. 5. *The multilayer perceptron*

MLPs are used to solve classification, prediction, and regression problems that require processing large amounts of data. They are used in financial forecasting, text analysis, and in medicine

for diagnosing and predicting diseases. One of the key features of MLPs is the use of nonlinear activation functions, such as ReLU, which allows the networks to learn more complex patterns.

Convolutional Neural Networks (CNNs) [31…34]. CNNs (Fig. 6) were developed to work with data that has spatial structure, such as images. In the mid-1980s, the first models were proposed that used convolution operations to extract important features in images. One of the first successful applications was the LeNet algorithm, developed by Yann Lekun for automatic recognition of handwritten digits.

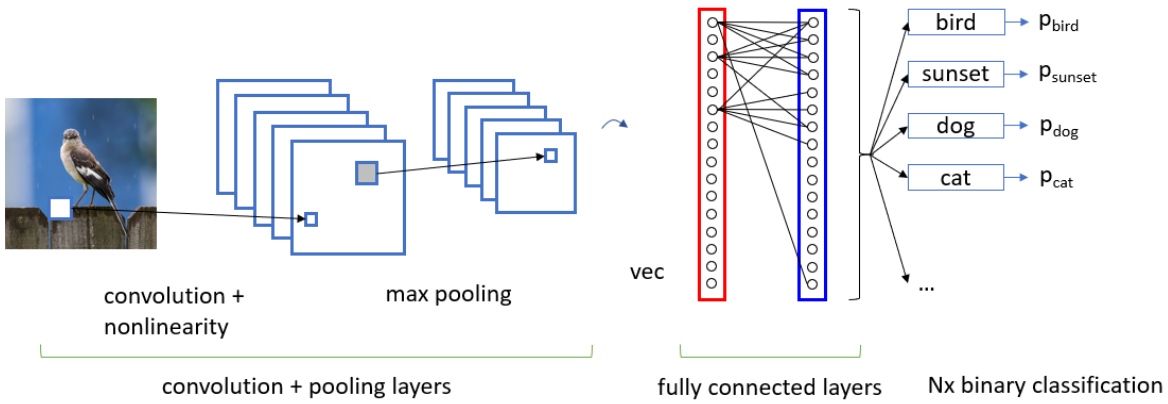

Fig. 6. *The convolutional neural network*

Imagine a photo: it's just a big matrix of numbers, where each number represents the brightness or color of a particular pixel. But to "see" the edges of an object or some special texture in this photo, for example, you need to pay attention not to individual pixels, but to their local groups.

This is where the convolutional layer comes into play. Simply put, it works like a small window or filter that "slides" (or, technically, performs a convolution) over the entire image, analyzing small areas at a time. This filter is also a matrix of numbers (a set of weights) that is learned during the training of the network.

Each step of such a "sliding" involves:

– multiplying the numbers in the filter and the corresponding numbers in the image;
– adding the results of the multiplication;
– writing the resulting number into a new feature map.

For example, let our window be 3×3 pixels (a matrix of size 3 by 3). It is applied to a specific part of the image, for example, a corner where there is a transition from light to dark color. If the filter is specifically configured to detect such a transition (border), then the result of the convolution will be a large number. If the filter is applied to a homogeneous area without changes, the result will be close to zero.

After the convolutional layers, a subsampling operation (pooling) is usually applied, which further reduces the dimensionality of the data and helps the network focus on the most important information.

Why is this important? Without convolutional layers, networks would have to process each pixel individually, which is inefficient and often does not allow the network to "understand" the

global structure of objects. Thanks to convolution, the network learns to see complex images, starting with simple details.

Thus, the convolutional layer extracts important local features as borders, corners, textures, and patterns, without paying attention to unnecessary information.

A key feature of CNNs is the presence of convolutional layers that allow for the automatic detection of important image features such as contours, textures, and objects. These networks are used for face recognition, image classification, automated driving, and in medical applications, where they are used to analyze medical images (MRI, X-rays).

Recurrent Neural Networks (RNNs) [35…37]. RNNs (Fig. 7) differ from traditional networks in that they have feedback, which allows them to store information about previous processing stages. This feature makes RNNs ideal for working with sequential data where context is important, for example, for text processing or time series analysis.

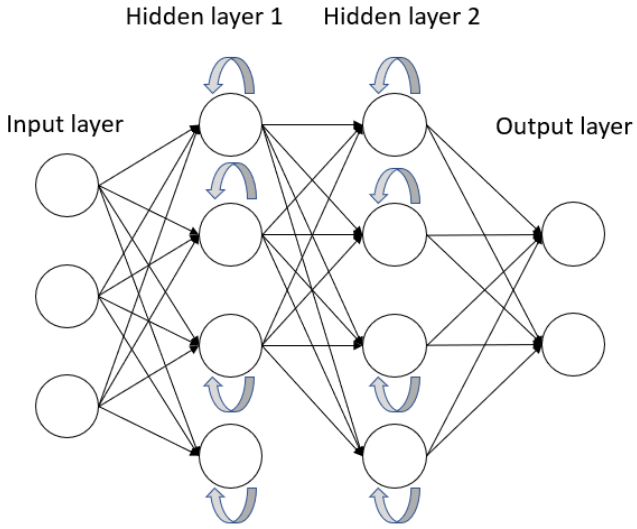

Fig. 7. *General form of RNN*

Let's suppose that we have a text and it is read word by word. Each new word makes sense only in the context of the previous ones. For example, the word “bank” can mean a financial institution or a riverbank, depending on what was said before.

RNN has a special mechanism: at each processing step, it receives:

– the current input (for example, a new word);

– an internal memory state (hidden state), which stores information about previous inputs.

After processing a new input element, the network updates its internal state and passes it on. Thus, the network seems to “flow memory”, which changes depending on the data flow.

Technically, at each step:

– the input x_t and the previous memory state h_{t-1} are combined;

– a new memory state h_t is calculated (usually through a nonlinear transformation);

– an output is formed (which may be the same or different from h_t, depending on the task).

Features of RNN:

1. Suitable for sequences of any length: texts, videos, time series, etc.

2. The idea of memory: the current output depends not only on the current input, but also on the entire previous context.

3. Ability to process data streams: suitable for tasks where data arrives gradually, for example, in chatbots.

The main variants of RNNs, such as LSTM (Long Short-Term Memory) and GRU (Gated Recurrent Unit), were developed to overcome the problem of fading gradients, which hindered efficient learning on long sequences. These networks are used in machine translation, text generation, speech data analysis, and speech recognition systems.

Generative Adversarial Networks (GANs) [38…40]. GANs (Fig. 8) were proposed by Ian Goodfellow in 2014 and quickly became popular due to their ability to generate new data that looks realistic. GANs consist of two main parts: a generator, which creates fake data, and a discriminator, which tries to distinguish between real and fake data.

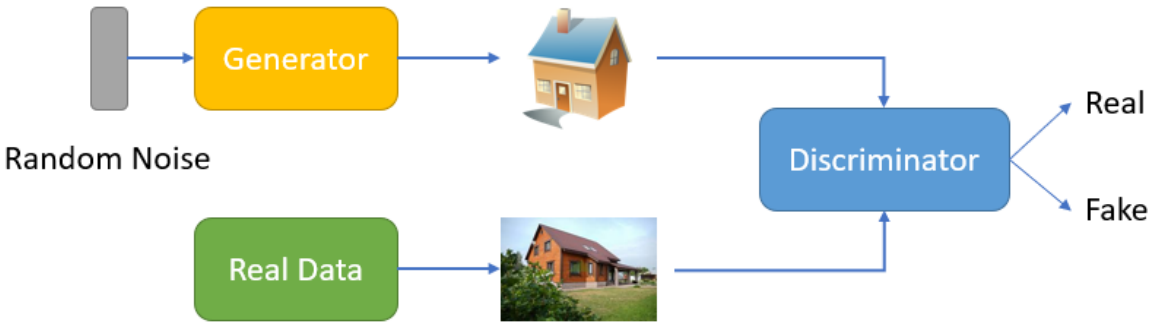

Fig. 8. *Structure of the GAN*

The idea of generative adversarial networks (GANs) has been one of the most groundbreaking in the field of artificial intelligence in recent decades. These networks have laid the foundation for a new paradigm in which artificial intelligence can not only analyze existing information, but also create fundamentally new content: images, videos, texts, and music by competing with two neural networks.

The GAN architecture is built on the interaction of two components: a generator and a discriminator. The generator tries to create data that looks like real data, while the discriminator learns to distinguish real data from artificially generated data. In the initial stages, the generator produces chaotic, unrealistic examples, but under the pressure of the discriminator, it gradually improves its ability to simulate the distribution of real data. The discriminator, in turn, becomes more and more demanding, trying to notice even the smallest deviations from reality.

This interaction resembles the process of an artist learning under the guidance of a harsh critic: each new attempt at drawing, each critical assessment contributes to the fact that the works become more and more convincing. Such a “competition” continues until the examples created by the generator become so realistic that the discriminator can no longer distinguish them with high confidence from the real ones.

Mathematically, this process is formalized as a minimax game: the generator minimizes the probability of being caught, and the discriminator maximizes the accuracy of recognizing fakes. Thanks to this idea, GANs have proven to be an extremely flexible tool for modeling complex data distributions without the need for explicit analytical formulas.

One of the features of generative networks is that a random vector (noise) is fed to the generator input, from which the network creates structured information. Gradually, it learns to transform this noise into objects that correspond to the distribution of the target data. For example, instead of a shapeless set of pixels, faces, natural landscapes, or handwritten numbers begin to appear, depending on the data on which the training is performed.

Issues of stability of training play a special role in the development of GANs. Sometimes one of the networks (generator or discriminator) far outweighs the other, which leads to the collapse of training. Therefore, various stabilization techniques are used in modern implementations: modifications of the loss function, alternative optimization procedures, and specialized architectural solutions.

Generative adversarial networks are already used in many areas today. In computer vision, GANs help create photorealistic images for video games or films, in medicine, to generate synthetic data for training diagnostic models, where real data is limited due to ethical considerations. At the same time, GANs find application in cybersecurity: for example, for modeling attacks and testing the stability of protective systems in environments where real attack data is limited or too sensitive for widespread use.

The hardware side of GAN implementation requires powerful computing resources. The large amount of parallel computing during the generation and evaluation of samples makes a GPU or a TPU a mandatory element of the infrastructure. In the inference process, optimization of the speed of work is of particular importance, since the models can be used in real time to generate personalized content or test security systems.

Thus, generative adversarial networks represent not just a technical solution, but a new philosophy of data generation and the interaction of artificial intelligence with the world. They have proven that it is possible to teach a machine to create and that in the future, this ability will become one of the determining factors of technological development.

GANs are used to generate photorealistic images, create videos, improve image quality, and even create music. One of the most famous implementations is StyleGAN, which allows you to create images of people who do not exist, but look completely realistic.

The Transformers [41] architecture has been one of the most important breakthroughs in AI, completely changing the way models can work with sequential data. First introduced in 2017 in the paper Attention Is All You Need by Google researchers, the concept offered an alternative to traditional recurrent neural networks (RNNs) and their more complex variations like LSTMs and GRUs.

The main innovation of Transformers is the principle of attention. In traditional networks, sequences were processed step by step, which limited the speed and efficiency when working with large amounts of data. Transformers offered a completely different approach: instead of linear processing, each element of the sequence can directly access all other elements through a self-attention mechanism. This allows the model to analyze the entire context at the same time, determining which parts of the input information are most important for the current task.

The self-attention mechanism works in such a way that each word in the input phrase is represented as a vector, and the transformer “asks” other words: which of them is important for understanding the current word? For example, in the sentence “Maria gave the book to Peter, because he loves to read” the transformer can understand through self-attention that the pronoun “he” refers to “Peter”, not “Maria”, even if these words are at a distance. Such deep capture of the context has become revolutionary for natural language processing.

The architecture of the transformer consists of two main parts: an encoder and a decoder. The encoder accepts input data and builds its internal representation, while the decoder generates the output result, a translation, an answer to a question, or some other form of information processing based on this representation. Thanks to parallel processing of sequences, transformers have significantly reduced the training time of large models and opened up the possibility of scaling to unprecedented sizes.

A separate role in the development of transformers was played by the technology of positional encoding, which allows the network to understand the order of words, since, unlike RNNs, transformers do not have a built-in concept of sequence. Using mathematical functions (usually sinusoidal), each element receives additional information about its position in the input phrase.

Since their appearance, transformers have become the basis for the most powerful artificial intelligence models. Well-known architectures such as BERT, GPT, T5, and others are modifications of the basic transformer, adapted to various tasks: from filling gaps in texts to generating new text blocks or even program code.

The practical results of using transformers have been impressive. Models based on them have first begun to overcome traditional boundaries in text comprehension tests, automatic translation, and dialogue construction. Due to the flexibility of the architecture, transformers have also been adapted for other types of data, for example, for image processing in Vision Transformer (ViT), where the image is divided into patches that are processed similarly to words in text.

The hardware implementation of transformers requires large computing power, especially for training models with hundreds of billions of parameters. Graphics processors, tensor processors, and specialized cluster solutions have become an integral part of the modern ecosystem for developing such models. Optimizations such as mixed precision training have also appeared to reduce memory consumption and processing time.

Transformers have moved from traditional text processing to a universal approach to work with information. Their ability to build complex dependencies in data and work in conditions of large-scale computing has made them the basis of a new era of artificial intelligence, an era where the creation and understanding of information are no longer limited by rigid rules, but develop in flexible, adaptive forms.

Transformers and their variants, such as BERT, GPT and T5 have shown incredible results in text processing tasks. They are used in applications such as machine translation, text generation, text classification, automatic translation, and even in code generation. Transformers allow training very large models on large data sets, which makes them extremely powerful.

Next, we consider two main principles of AI, machine learning and deep learning.

Machine learning and deep learning [42…44] are not just abstract terms, but technologies that are already changing our daily lives. They allow computers to “learn” from data, to perform complex tasks that were previously impossible to automate, and to do so without direct human intervention. Today, machine learning has become the basis for a multitude of technological achievements, from recommendation systems in online stores to autonomous cars. However, these technologies, while revolutionary, have their characteristics, advantages, and limitations. Machine learning (ML) is a powerful tool for analyzing large amounts of data and making decisions based on this data. Imagine that a system needs to learn to recognize photos of cats and dogs. The classic approach would be for a programmer to manually identify key features of each animal, such as the shape of the ears or characteristic skin features. But with machine learning, everything changes: instead of specifying all these features, we simply provide the system with a large number of

images in which cats and dogs are clearly labeled. The model then "trains" itself to find common features for each class based on this data. This allows it to classify new images quickly and accurately, without having to manually specify all the details. Machine learning works on the principle of feedback learning. When the system makes mistakes, it adjusts its assumptions, improving the accuracy of its predictions with each step. This is similar to the process of human learning, where we gradually improve our skills by observing the results of our actions. However, one of the main advantages of machine learning is that it can work with much larger amounts of data than humans and detect complex patterns that may not be obvious to us.

If machine learning is a good tool for processing data, then deep learning is its evolution, which uses neural networks with many layers to achieve much more complex results. The basic idea of deep neural networks is that they consist of several layers of neurons, each of which has its role in the information processing process. The more layers, the more complex and abstract the features that the system can detect in the data.

For example, one of the most common architectures used for image analysis is CNN. While simpler models can only detect basic features, such as edges or textures in an image, in a CNN, the neurons will be able to automatically detect increasingly complex elements. For example, in the first layer, the network can focus on simple geometric shapes, such as lines and angles. Then, in deeper layers, it will detect more complex elements, such as facial contours or specific objects in the image.

This allows systems using CNNs to make more complex predictions, such as diagnosing diseases based on medical images. In medicine, CNNs are already used to analyze X-rays, MRIs, and even histological samples to detect cancer cells. These technologies have the potential to surpass the accuracy of human doctors because they can process much larger volumes of images and take into account more factors.

Deep learning also includes recurrent neural networks (RNNs), which are used to work with sequential data, such as text or speech signals. Traditional neural networks usually work with fixed data sets, but RNNs can process sequences while preserving the context of information from previous stages. For example, in natural language processing tasks, this allows a model to "remember" previous words or even the context of an entire phrase to correctly predict the next word or generate more natural text.

These networks find applications in tasks such as machine translation, automatic subtitling, or chatbots. However, recurrent networks have their limitations, and special modifications such as Long Short-Term Memory (LSTM) have been developed to solve some problems. LSTMs are able to store important information in the long term, which is critical for understanding more complex or longer texts where temporal context is important.

Machine learning and deep learning have great benefits in many industries. One of the main advantages is the ability to analyze huge amounts of data, doing it faster and more accurately than a human. Thanks to this, these technologies have become indispensable in medicine, finance, industry, and even in the development of smart cities. However, not all aspects of these technologies are perfect. One of the main limitations is the need for large amounts of data for training. If the data is insufficient or of poor quality, the model may not be as accurate as we would like. In addition, even the most advanced models can be a black box, where it is difficult to understand how the system came to a certain conclusion. This becomes a problem in critical applications, such as medicine or law enforcement, where transparency and explainability of decisions are required.

Another limitation is related to computational resources. Training large deep learning models requires significant power, including GPUs and specialized hardware. This makes access to these technologies more expensive and limits their use for many companies or organizations with limited budgets.

Machine learning and deep learning have not only become the basis for the development of many new technologies, but also significantly changed the approach to solving complex problems. They allow you to process huge amounts of data, automate numerous processes, and improve the accuracy of predictions and decision-making. However, with these capabilities come certain limitations that require attention when developing and applying these technologies. The lack of transparency in the operation of models, the need for big data, and significant computational resources are factors that must be taken into account to make the technologies safe, effective, and accessible to everyone.

Part 3. Explainable AI

Artificial intelligence is rapidly penetrating our lives, shaping decisions in areas where human responsibility used to be the main one: in medical diagnoses, credit decisions, court cases, and even in the selection of news that we read every day. However, along with this development, an important question arises: can we understand why a machine decided this way and not another? How can we be sure that its decision is fair and logical?

The basis of this problem is that many modern algorithms work like so-called "black boxes". Take, for example, a deep neural network that makes credit decisions. To an outside observer, this system looks like a mysterious mechanism: data about the applicant is fed into the input, and the output is approval or rejection. But the criteria by which the decision was made, which influenced him the most, remain hidden. This creates the risk of bias, injustice, and even legal irresponsibility.

To overcome this danger, a whole field of research has emerged under the general name of Explainable Artificial Intelligence (XAI) [45…47]. The goal of XAI is to return control of decision-making processes to humans: to make the machine not only give an answer, but also explain the logic of its thinking.

The XAI model is shown in Fig. 9.

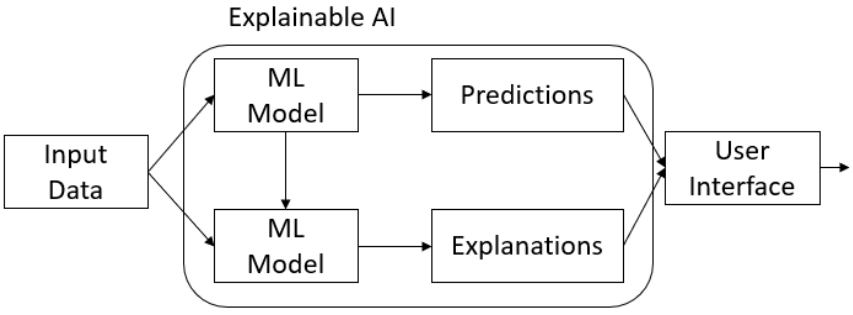

Fig. 9. *Schematic representation of the XAI model*

One approach is the idea of local interpretation. For example, if a neural network classifies a photo as a "dog" we can create dozens of modified versions of that photo, removing certain areas, and see if the answer changes. If hiding the ears or muzzle significantly changes the prediction, it

means that these features are key to the classification. In this way, even a complex model becomes a little closer to human understanding.

Another approach is based on the ideas of cooperative game theory. Imagine a group of players who have won a large sum of money together, and you need to fairly distribute the winnings according to each person's contribution. Similarly, algorithms can assess how much each data feature "added" or "subtracted" from the decision. For example, for a loan applicant, it might look like this: his high salary positively influenced the decision, his young age slightly reduced the chances, and the absence of debt added a few important points. With such detail, the algorithm's decision becomes not only acceptable, but also understandable and logical for a person.

Particularly interesting explanation techniques have emerged in natural language processing. When we ask large language models like GPT or BERT questions, they don't just give us an answer. Their internal mechanisms, the attention mechanisms, allow us to see which words or phrases they paid the most attention to while searching for the right answer. If we ask, "Who was the first president of Ukraine?", the model focuses on the words "first" and "Ukraine," telling us that it is truly analyzing the content of the question, and not looking for a random answer.

Explainable AI becomes especially critical in areas where human life or fate depends on the correctness of decisions. In medicine, image analysis systems must not only indicate a diagnosis, but also explain what features on an MRI scan were the basis for such a conclusion. In banks, it is necessary that customers understand the reasons for the refusal of credit; otherwise, trust in the financial system is destroyed. In courts, decisions of crime prediction algorithms without the possibility of verification and appeal can become a source of systemic injustice.

Therefore, algorithmic transparency is not just a fashion trend but a fundamental requirement for responsible technological development. Only when we teach machines to explain their actions can we build a true partnership with them, a partnership where trust is based on understanding.

The real breakthrough in the topic of algorithmic transparency came with the advent of modern language models like GPT-4. These models demonstrate phenomenal abilities, they can maintain a dialogue, write code, translate texts, generate poetry, and analyze legal documents. But with this power comes a new responsibility. The user should be able to ask: why did the model give this answer and not another? And that is why companies that develop such systems are increasingly integrating elements of XAI into their products.

One solution was the introduction of tools for "explanation of the logic of the answer". For example, if GPT-4 provides legal advice to a user, the system can clarify: "This decision is based on a precedent from such and such a year", or: "The answer is formed based on the most common approaches in open sources". Such a capability not only builds trust but also allows the user to understand more deeply how the model functions.

Similar ideas are used in the field of computer vision. For example, when a model trained on millions of images analyzes a chest X-ray, it doesn't just give a verdict like "possible lung damage". It highlights the areas that influenced that decision in color. This way, the doctor can see exactly where the system "saw" pathology and can check its conclusions without having to trust the machine blindly. This is especially important when it comes to early cancer detection or complex pulmonary cases. Another fascinating example is generative image models, such as DALL·E, which create pictures based on text descriptions. While such models seem almost magical, there is also an explanation for how they work. They are based on multi-level analysis of meaning: for example, if a user types "cat in a hat in the rain in the style of Van Gogh", the model doesn't just compile an image from those words. It sequentially models the concepts of "cat",

"hat", "rain", and then imposes a style that it is trained to recognize as "Van Gogh". And each of these steps can be interpreted by studying which examples from the training set influenced this result, or which vectors in the feature spaces were activated.

XAI developers have also begun to implement new interfaces that allow not only to see the final result, but also to "peek inside" the process. For example, in dialogues with large language models, functions such as "ask about reasoning" or "show the steps that led to the answer" appear. This not only creates transparency but also turns the user from a passive consumer into an active participant in the decision-making process.

And even more interesting are attempts to teach the model self-analysis. Some new modifications of GPT allow the model to "rethink" the answer and find its own mistakes. For example, a user might ask, "Rate the correctness of the previous answer and explain why there might have been an error". This opens the door to deeper systems that can not only respond but also reflect, just like a human.

Transparency in AI is, in essence, the language of trust. In a world where technology increasingly influences daily choices, we need to understand how and why a particular algorithm works. And while no model can be completely "understandable" to humans, the desire for explainability is the compass that guides us to the responsible and fair use of artificial intelligence.

Recommender systems have become one of the key elements of the digital ecosystem, as they provide users with access to personalized content, products, or services. From a scientific point of view, recommender systems are complex information and analytical systems based on machine learning methods, big data processing, and optimization algorithms.

Let us consider the types of modern recommender systems [49, 50]:

1. Collaborative filtering is based on the assumption that users with similar preferences in the past will show similar interests in the future. The classic approach to collaborative filtering involves constructing a matrix of users and objects (items), where the elements reflect the level of interaction or evaluation. Since such matrices are usually very sparse, dimensionality reduction methods are used to process them, among which the singular matrix decomposition (SVD) occupies a special place. When using SVD, a large sparse matrix is decomposed into several matrices of lower dimension, which allows you to identify latent factors of interaction between users and objects. Thus, instead of working directly with a small number of reviews, the system models the hidden tastes and characteristics of objects, making predictions based on vector representations in the latent space. Example of application: the Netflix Prize Challenge algorithm used an extended version of matrix factorization to improve the quality of movie recommendations, which made it possible to significantly reduce the error in predicting ratings.

2. Content filtering is based on the characteristics of the objects themselves and the user profile. The goal of the algorithm is to find objects that most closely match the user's known interests based on a set of features. For example, if the user often reads science fiction, the system will recommend new works in this genre, taking into account metadata such as genre, author, subject matter, and writing style. Technically, the user profile and the object profile are represented as feature vectors. The similarity between them is calculated using proximity metrics (e.g., cosine similarity). Such models work well in cases where a large amount of structured information about objects is available.

3. Hybrid systems. The hybridization of recommendation approaches has become a response to the limitations of both collaborative and content filtering. Hybrid systems combine the advantages of both methods by integrating them into a single model. The most common

hybridization methods include cascading models, a weighted combination of their outputs, or building ensembles at the latent feature level. An example of an application: modern streaming service systems such as Spotify or Netflix use hybrid architectures, where collaborative filtering identifies a group of potentially interesting objects, and content filtering refines their rating according to the user context.

4. Deep neural networks in recommendation systems. The development of deep learning has radically changed approaches to recommendations. Instead of manually forming user and object profiles, modern systems use deep neural networks (DNNs) to automatically extract hidden patterns from data. Frameworks like Wide&Deep, developed by Google for Play Store Recommendations, combine linear models and deep networks for an optimal balance between memorization of known patterns and generalization of new trends. More advanced models, such as DeepFM or DIN (Deep Interest Network), use attention mechanisms to study in detail the interaction between the user's behavior history and new objects. Such systems can model not only general preferences but also dynamic interests in real time.

5. Recently, graph neural networks (GNN) have become popular in recommendation systems. In such approaches, data is represented in the form of a graph, where nodes are users and objects, and edges are interactions between them. GNNs are trained directly on the structure of the graph, revealing complex relationships that are inaccessible to traditional models. Application example: TikTok uses graph models to personalize the video feed, taking into account not only the user's behavior, but also the behavior of their "similars" in the context of the content.

6. Transformer-based models. Transformer-based models (e.g., SASRec, BERT4Rec) are implemented to model sequences of user interactions. Thanks to self-attention mechanisms, transformers can determine which previous actions have the greatest impact on the user's next choice. This allows you to take into account complex, nonlinear patterns in behavior, for example, to predict interests depending on the context of recent views or the dynamics of changes in user preference.

Modern recommendation systems, especially those based on deep learning, work with extremely large amounts of data and complex models containing millions or even billions of parameters. Effective training and deployment of such systems requires consideration of both algorithmic and hardware features.

The process of training modern recommendation models, such as DeepFM, DIN, SASRec, or BERT4Rec, involves processing gigantic datasets and computing a huge number of gradients to optimize the network weights. Traditional CPUs, due to their architecture focused on sequential processing of tasks, are not able to provide the necessary level of parallelism.

After the training phase is completed, the models must work in real time, providing recommendations in minimal time. The main challenge here is not only accuracy, but also inference latency, i.e., the time it takes to issue a result after receiving a request.

To achieve low latency, specialized frameworks for optimizing neural networks are widely used:

– TensorRT (developed by NVIDIA) allows you to convert neural networks into a form that is most efficient for inference on GPUs. TensorRT automatically applies optimizations such as layer fusion (combining adjacent operations), quantization of models in INT8 or FP16, and structural optimization of the computation graph.

– ONNX Runtime provides a platform-independent mechanism for accelerating models that are stored in the Open Neural Network Exchange (ONNX) format. ONNX Runtime can run

on various computing platforms, including CPUs, GPUs, and specialized accelerators, using internal optimizations to increase throughput and reduce latency.

Thus, the real success of a recommender system today depends not only on the quality of the model but also on its engineering optimization for latency and computational requirements.

Traditionally, model inference was performed on central server computing clusters. However, with increasing demands for speed, privacy, and scalability, there has been a need to move processing closer to the user, directly to the devices they use. This concept is known as edge inference.

In edge inference systems, models are optimized to run on mobile phones, tablets, consumer devices, or specialized edge servers. As a result, several key benefits are achieved:

– Reduced latency. Since requests are processed locally, there is no need to transmit data over the network to a remote data center and back, critical for real-time applications (for example, recommendations in the news feed or personalized advertising).

– Increased data privacy. User data can be processed directly on the device, reducing the risk of leakage or misuse.

– Optimization of network resource usage. Since most of the processing is performed locally, the load on the network infrastructure is reduced.

Technically, edge inference requires a significant reduction in the size of models without a significant loss of accuracy. For this, the following methods are used:

– Quantization. Transferring model weights and activations to a lower bit size (for example, INT8 instead of FP32).

– Thinning and pruning, removing insignificant weights and connections in a neural network.

– Using lightweight architectures, such as MobileNet, EfficientNet-Lite or TinyBERT for tasks that require minimal resources.

Example of practical use: Google Assistant on mobile devices uses edge inference to provide instant recognition of voice commands even without Internet access. In recommendation systems, in particular on mobile platforms YouTube or Instagram, miniaturized models are used to predict content, which increases personalization and at the same time protects user privacy.

Part 4. Artificial intelligence in cybersecurity: Future Technologies to protect today

In today's world of rapid digitalization, cybersecurity is becoming a critically important component of protecting personal, corporate, and government information. The growth of data volumes, the complexity of cyberattacks, and the constant improvement of criminals' methods require fundamentally new approaches to protecting information systems. AI opens a new era in cybersecurity, allowing the creation of adaptive, self-learning, and proactive defense tools. From detecting anomalies in network traffic to intelligent analysis of events, from the latest encryption methods to steganographic systems, AI is becoming a powerful tool for protecting our digital reality [51…53]. We will consider the prospects for the development and implementation of artificial intelligence technologies in key areas of cybersecurity.

Artificial intelligence is rapidly changing the cybersecurity landscape today, offering new methods of protection in the era of digitalization of all spheres of life. While traditional security systems have relied primarily on a set of predefined rules and signatures of known attacks, AI-based approaches allow for the creation of dynamic defense mechanisms that learn and adapt to new, as yet unknown threats.

One of the key advantages of using AI in cybersecurity is its ability to work with large amounts of data in real time. Modern digital systems generate terabytes of data every day, from activity logs to network traffic flows. A person is physically unable to process such volumes without the help of automated systems. Machine learning algorithms, in turn, can quickly detect anomalous patterns that may indicate malicious activity or data leaks.

A promising direction is the implementation of deep learning to analyze complex attacks that do not have clear signs and have been secretly deployed in the system for a long time. For example, unsupervised learning methods allow you to detect unknown types of attacks that do not have corresponding signatures in databases by analyzing atypical behavior of users or the system.

Another important trend is the development of predictive threat analytics. Using predictive models built based on historical incident data, systems can not only respond to attacks after they are detected, but also predict them at the preparation stage. This changes the very concept of protection, from a reactive approach to a proactive one.

The role of artificial intelligence in ensuring data confidentiality deserves special attention. Private machine learning methods (for example, differential privacy, federated learning) allow you to create systems that learn on distributed data without its direct transmission, which significantly increases the level of protection of personal information.

It is also important to note that AI has the potential to strengthen cryptographic methods [54]. The use of generative models to create new cryptographic algorithms or the automation of processes for checking their stability opens up new areas of research in the field of data security.

In the context of the future of digital security, it is important to realize that AI is becoming not only a protection tool, but also an object of threats itself. Therefore, along with the implementation of intelligent systems, it is necessary to develop mechanisms for their protection against manipulation, attacks on training data (data poisoning), and abuse.

Thus, the use of artificial intelligence in cybersecurity is not just an expansion of existing capabilities but a fundamental transformation of the entire architecture of information environment protection, which has the potential to ensure a secure future in the face of the continuous evolution of digital threats.

In the modern digital environment, attacks are becoming increasingly sophisticated, often avoiding detection using traditional methods based on signatures or simple heuristic rules. This necessitates the use of more flexible and intelligent systems capable of identifying both known and unknown threats. Artificial intelligence, in particular machine and deep learning methods, has opened up new horizons in this area.

Unlike classic intrusion detection systems (IDS), which focus on detecting specific signs of already known attacks, AI approaches can build profiles of "normal" network, user, or system behavior and automatically detect deviations from these profiles. This is where unsupervised learning algorithms, such as clustering and data dimension reduction methods, come to the fore. They allow you to detect anomalies even without prior knowledge of what a potential attack looks like.

A specific example is autoencoders, which learn to compress normal data and reconstruct it with minimal loss. If an anomalous or malicious message is fed to the input, the autoencoder demonstrates a high reconstruction error, signaling a possible attack.

Reinforcement learning methods are also increasingly used in attack detection systems, where agents learn how best to respond to suspicious activity, receiving a reward for correctly detecting or preventing intrusions. This allows for systems that not only passively observe but also actively intervene to neutralize threats.

Deep neural networks, including recurrent (RNN, LSTM) and transformative architectures, have proven their effectiveness in analyzing sequences of events in network logs, system logs, or traffic flows. They are able to capture long-term dependencies in data and detect complex attacks that unfold over a significant period of time, such as so-called "slow data exfiltration" attacks.

In addition to detecting attacks in real time, artificial intelligence allows you to automate the process of correlating incidents and forming responses to them. The integration of AI into SIEM (Security Information and Event Management) platforms transforms them from passive log collection systems into active cyber defense tools that can independently assess risks and generate recommendations for operators.

However, the implementation of AI in attack detection also has its challenges. One of the main problems is the problem of false positives, which can overwhelm security analysts. To solve it, hybrid models are being developed that combine the capabilities of deep learning with expert systems, balancing between automaticity and controllability of the process.

As a result, artificial intelligence is changing the fundamental idea of attack detection systems: from passive sentinels that react to known threats, they are transformed into active defenders capable of predicting, detecting, and adapting to new forms of cyber threats. This is especially important in a world where technology is developing at lightning speed and the number of potential attack vectors is constantly growing.

In the world of digital security, event logs and network traffic are key sources of information about the state of systems, user activity, and possible signs of malicious activity. However, the volume of this data in modern organizations is gigantic; terabytes of logs and network records are generated every day. Traditional analysis approaches based on static rules or manual processing have proven insufficient to detect complex or previously unknown threats in such an environment. This is where intelligent systems based on artificial intelligence come into play. The main idea of intelligent analysis of logs and traffic is to move from simple pattern search to a deep understanding of the context of events. Artificial neural networks, especially transformer-based architectures, trained on large volumes of logs, can detect hidden relationships between events that at first glance seem unrelated. For example, unusual account activation outside of business hours, combined with unusual requests to internal servers, can be an early indicator of compromise.

As it was told, to work with real-time events, deep recurrent models (RNN, LSTM) are used, which are able to analyze sequences of events and build dynamic profiles of system and user behavior. This allows you to instantly capture anomalies that could go unnoticed in traditional systems: for example, a gradual increase in the number of connections to a particular server, which precedes a denial-of-service attack (DDoS).

In addition, deep cluster analysis algorithms are used to automatically classify events according to their likely risk. For example, systems based on unsupervised learning algorithms are able to group events by similar behavioral characteristics and highlight atypical groups for further investigation.

Another important technology is the use of graph neural networks (GNN), which build structures of relationships between objects: accounts, IP addresses, and devices. Analysis of such graphs allows you to detect complex attack scenarios, such as lateral movement within a corporate network, where an attacker gradually moves from one compromised object to another.

One of the advantages of intelligent systems is the ability to self-learn. The use of Active Learning technologies allows systems to interact with analysts: when the system doubts the

classification of an event, it can initiate a request for human verification and improve its models accordingly in the process. This creates a continuous cycle of increasing detection accuracy.

At the same time, the problem of explainability of analysis results remains important: operators must understand why a certain event was classified as suspicious. Therefore, XAI systems for log analysis are increasingly being developed, which allow not only to detect a threat, but also to justify their decision, for example, by building cause-and-effect relationships between user actions.

Intelligent log and traffic analysis systems are becoming the basis of modern incident response centers (CIRC, computer incident response center), transforming the monitoring process from passive observation to proactive threat search and prevention of cyberattacks at the early stages. In the future, it is such technologies that will ensure the stability of digital infrastructure in a constantly changing world.

Cryptography, as the art of information protection, has always gone hand in hand with the development of mathematics and computing technologies. However, with the advent of artificial intelligence, the cryptographic industry has entered a qualitatively new era. Traditional encryption and decryption methods, which were based mainly on deterministic algorithms, are beginning to be supplemented with adaptive and self-learning mechanisms, which opens up new horizons for data security.

One of the most exciting areas is the use of neural networks for the automatic generation of cryptographic schemes. For example, deep learning can be used to create encryption systems that are optimized not according to set rules, but by learning the interaction between a conditional "sender" and "receiver" of messages. Such approaches have been researched in experimental work: neural networks learned not only to encrypt data, but also to build their strategies for resisting decryption attacks. The encrypting network and the cracking network competed with each other, constantly improving their methods, which led to the creation of non-standard, difficult-to-predict ciphers.

Another extremely promising area is the use of artificial intelligence to assess the stability of cryptographic systems. Instead of manually analyzing an algorithm or relying on classic attacks, neural models can uncover hidden vulnerabilities by using machine learning to model potential ways of compromise. For example, using reinforcement learning techniques allows you to create agents that independently examine a cryptographic protocol in search of weaknesses, which often remain inaccessible to traditional testing.

AI is also changing approaches to key management. The distribution and storage of cryptographic keys is one of the most vulnerable points of any security system. Using intelligent agents to monitor the state of the key infrastructure, predict possible leaks, and automatically redistribute keys significantly improves overall security.

One particularly interesting idea is the development of the concept of "adaptive cryptography", where encryption parameters change in real time under the influence of behavioral models. For example, in the case of detecting anomalous activity of a user or device, the system can automatically increase the level of encryption, change the algorithms used, or even completely update cryptographic keys without human intervention.

However, it is worth noting that AI is not only a powerful tool for protection, but also a potential threat to traditional cryptography. In particular, with the development of quantum computing and combined neuroquantum models, there are risks of revealing classical cryptographic schemes much faster than previously expected. Therefore, much attention is paid

today to the development of post-quantum cryptographic algorithms, which are also beginning to integrate artificial intelligence methods to increase their flexibility and resilience.

Thus, the synergy of cryptography and artificial intelligence forms a new battlefield and new opportunities at the same time. In the coming years, it is the integration of AI into cryptographic technologies that will determine the level of security in critical areas, from the protection of government systems to the security of personal communications in the global digital network.

Steganography, the science of hiding the very presence of information, is experiencing a real rebirth thanks to the capabilities of artificial intelligence [55]. While classical methods focused on manipulating individual bits of digital objects: images, audio, or texts, today neural networks allow us to create information hiding at fundamentally new levels, using the properties of deep data representations.

One of the most exciting areas is the use of generative models, in particular generative adversarial networks (GANs), for steganography. Here, the hidden information is integrated into the object so naturally that even specialized steganoanalytic methods lose their ability to detect it. For example, using GANs, you can train the network to create realistic images in which the message "dissolves" in textures, colors, and microstructures without violating the visual integrity of the object.

A feature of the approach based on artificial intelligence is the optimization of the balance between concealment and robustness. The neural network can automatically adapt the embedding process to the characteristics of each medium: in the area of high image variability, more data can be hidden, and less data in the area of uniform shades, minimizing the risk of detection. This level of adaptability is almost unattainable in traditional steganographic systems.

No less important is the use of AI in steganalysis, the process of detecting hidden messages. Classical steganography analyzers were usually based on the manual design of features: image statistics, spectral characteristics, etc. Today, deep neural networks independently learn to detect anomalies that humans are unable to intuitively describe. For example, convolutional neural networks (CNN), specially adapted to process small deviations, can detect traces of steganographic embedding even when it does not cause noticeable changes in image statistics.

Also, intriguing is the approach using attention models, "attention" mechanisms that allow networks to focus on potentially suspicious areas of the medium. Thus, the system does not analyze the image evenly, but pays more attention to areas where the probability of hiding information is higher, which significantly increases the efficiency of steganography.

In addition, artificial intelligence opens up the possibility of building fully autonomous steganographic systems, where the embedding module and the message extraction module co-evolve together. This means that the process of hiding and extracting information automatically adapts to changes in the data transmission environment: for example, in the event of a change in the JPEG compression ratio or the appearance of artifacts during transmission over unstable channels.

However, the development of AI in steganography and steganoanalysis also creates new challenges. On the one hand, artificial intelligence allows data to be hidden more and more imperceptibly; on the other, it equips analysts with powerful tools for their detection. This creates a kind of arms race in digital silence, where victory is determined not only by the skill of creating a medium, but also by the ability of artificial intelligence to see the invisible.

Thus, the integration of artificial intelligence into steganography and steganoanalysis does not simply expand the technical capabilities of these industries. It changes the very nature of digital

information hiding, turning it into a dynamic process of constant adaptation and confrontation, which is crucial for cybersecurity in a world of constant information turbulence.

In a world where the volume of data and the complexity of cyber threats are growing exponentially, traditional protections are increasingly proving to be insufficient. A human operator simply does not have time to analyze every alarm signal, recognize all possible attack vectors, or respond to threats in real time. This is where intelligent agents come to the fore, autonomous systems based on artificial intelligence that can act independently, learn, and make decisions without direct human intervention.

Intelligent agents in cybersecurity can be imagined as digital security guards who do not simply follow predefined rules, but also study the environment themselves, look for anomalies, predict possible threats, and develop appropriate countermeasures. They constantly learn from data, analyzing event logs, network traffic, user behavioral patterns, and even their own experience of previous attacks.

The technological basis for such agents is the methods of deep learning, reinforcement learning, and self-organizing systems. For example, an agent can receive "rewards" for timely detection and blocking of attacks or for minimizing damage in the event of an incident. Over time, through many cycles of trial and error, it forms response strategies that outperform static algorithms.

One practical embodiment of intelligent agents is Active Cyber Defense systems. Unlike traditional passive monitoring, active defense assumes that the agent can independently change routing rules, block suspicious IP addresses, isolate infected network segments, or even initiate countermeasures to gather intelligence about the attacker.

Another promising direction is multi-agent systems, where a large number of agents cooperate, exchange information, and divide tasks to more effectively protect large and complex infrastructures. For example, individual agents can specialize in different types of traffic: analyze HTTP requests, e-mail, and internal data traffic in the data center. By coordinating their actions, they provide deep, multi-layered security.

The use of intelligent agents is especially important to combat zero-day attacks, where traditional signature-based detection methods are ineffective. An agent can notice even minor deviations in system behavior that signal an unknown type of attack and trigger preventive defense mechanisms before the threat reaches catastrophic proportions.

However, the emergence of autonomous cyber defenders also brings with it new challenges. The question of trust arises: can we fully rely on the decisions of a system that is able to act independently? How to ensure that the agent will not be vulnerable to manipulation or training on previously prepared data (attacks on the learning process, poisoning attacks)? Therefore, modern researchers pay significant attention to the issues of transparency, auditing of agent decisions, and the security of AI systems themselves.

Intelligent agents in cyber defense are becoming more than just a useful tool, they are turning into an obligatory component of any defense strategy in the digital age. Their ability to act quickly, adaptively, and autonomously can become a decisive factor in the battle for cybersecurity in a world of constant change and invisible threats.

The modern world is developing at an unprecedented speed. Every new technology brings not only opportunities but also risks, and cybersecurity is the fine line that separates a stable digital world from chaos. In this dynamic reality, artificial intelligence ceases to be just a tool: it becomes the basis of a new security architecture.

The application of AI in attack detection, event and traffic analysis, cryptography, steganography, and building intelligent agents is fundamentally changing the very concept of protection. Now, cybersecurity is not just about defense; it is about staying ahead of threats, about actively studying the environment, about understanding the dynamics of the behavior of both users and potential attackers.

Intelligent systems are able to see what is hidden from the human eye: detect zero-day attacks, adapt their strategies in real time, and self-learn in new conditions. Moreover, they are able to think in terms of probabilities, risks, and optimal scenarios, which opens up new horizons for preventive protection.

However, this new world requires new responsibilities. The interpretability of solutions, the protection of learning processes, and the ethical aspects of using AI in security are becoming as important as the effectiveness of the systems themselves. Building trust in intelligent solutions is a challenge that must be solved together with their integration.

Artificial intelligence is not a replacement for humans in the field of cybersecurity. It is their ally, assistant, and outpost in the fight for a secure future. Proper use of its capabilities will allow not only to protect the present, but also to confidently design the future, a world where innovation and security go hand in hand.

Conclusions

The performed research highlights the profound evolution of artificial intelligence from early theoretical concepts to advanced, scalable technological systems that shape modern digital society. Historical analysis shows that AI development has been driven by critical scientific breakthroughs, the advancement of computational power, and the persistent ambition to create machines capable of learning and adaptation.

Technical foundations such as neural networks, deep learning architectures, Transformers, and GANs have enabled artificial intelligence to process complex data, recognize patterns, and generate new information with remarkable accuracy and flexibility. These technologies form the core mechanisms by which AI systems operate and continuously improve their capabilities.

At the same time, the complexity of modern AI models has created an urgent need for transparency and interpretability, leading to the rise of XAI as a fundamental research and development direction. Ensuring that AI decisions are understandable and trustworthy remains critical for the responsible integration of AI into high-stakes domains.

Finally, the integration of AI into cybersecurity demonstrates its transformative potential in protecting the digital ecosystem. Intelligent threat detection, anomaly analysis, cryptographic innovations, steganographic techniques, and the development of autonomous defense agents collectively represent the future of adaptive and resilient cybersecurity infrastructures. AI is no longer a distant technological aspiration but a foundational element for proactive digital security in an increasingly complex and interconnected world.

Thus, artificial intelligence stands today not only as a field of innovation but as a strategic cornerstone for building secure, transparent, and adaptive technological systems of the future.

REFERENCES

1. Karnouskos, S. (2022). Symbiosis with artificial intelligence via the prism of law, robots, and society. *Artificial Intelligence and Law*, 30(1), 93-115. DOI: 10.1007/s10506-021-09289-1.
2. Damar, M. et al. (2024). Super AI, Generative AI, Narrow AI and Chatbots: An Assessment of Artificial Intelligence Technologies for The Public Sector and Public Administration. *Journal of AI*, 8(1), 83-106. DOI: 10.61969/jai.1512906.
3. Cao, L. (2022). Ai in finance: challenges, techniques, and opportunities. *ACM Computing Surveys (CSUR)*, 55(3), 1-38. DOI: 10.1145/3502289.
4. Holmes, J. et al. (2004). Artificial intelligence in medicine. *Ann R Coll Surg Engl*, 86, 334-8. DOI: 10.1308/147870804290.
5. Zhai, X. et al. (2021). A Review of Artificial Intelligence (AI) in Education from 2010 to 2020. *Complexity*, 2021(1), 8812542. DOI: 10.1155/2021/8812542
6. Hu, Y. et al. (2021). Artificial intelligence security: Threats and countermeasures. *ACM Computing Surveys (CSUR)*, 55(1), 1-36. DOI: 10.1145/3487890.
7. Bertino, E. et al. (2021). AI for Security and Security for AI. *Proceedings of the Eleventh ACM Conference on Data and Application Security and Privacy*, 333-334. DOI: 10.1145/3422337.3450357.
8. Elliott, D. & Soifer, E. (2022). AI technologies, privacy, and security. *Frontiers in Artificial Intelligence*, 5, 826737. DOI: 10.3389/frai.2022.826737.
9. Jeong, D. (2020). Artificial intelligence security threat, crime, and forensics: taxonomy and open issues. *IEEE Access*, 8, 184560-184574. DOI: 10.1109/ACCESS.2020.3029280.
10. Atkinson, K., Bench-Capon, T. & Bollegala, D. (2020). Explanation in AI and law: Past, present and future. *Artificial Intelligence*, 289, P. 103387. DOI: 10.1016/j.artint.2020.103387.
11. Schaller, R. R. (1997). Moore's law: past, present and future. *IEEE spectrum*, 34(6), 52-59. DOI: 10.1109/6.591665.
12. Bell, G. (2014). Moore's Law evolved the PC industry; Bell's Law disrupted it with players, phones, and tablets: New Platforms, tools, and sevices. Microsoft Research, San Francisco, CA, USA, Technical Report, MSR-TR-2014-2.
13. Gokhale, P., Bhat, O. & Bhat, S. (2018). Introduction to IOT. *International Advanced Research Journal in Science, Engineering and Technology*, 5(1), 41-44. DOI: 10.17148/IARJSET.2018.517.
14. Akman, V. & Blackburn, P. (2000). Alan Turing and artificial intelligence. *Journal of Logic, Language, and Information*, 391-395. DOI: 10.1023/A:1008389623883.
15. Turing, A.M. (1950). Computing machinery and intelligence. *Computers & Thought*, 11-35.
16. Toosi, A. et al. (2021). A brief history of AI: how to prevent another winter (a critical review). *PET clinics*, 16(4), 449-469. DOI: 10.1016/j.cpet.2021.07.001.
17. Van Melle, W. (1978). MYCIN: a knowledge-based consultation program for infectious disease diagnosis. *International journal of man-machine studies*, 10(3), 313-322. DOI: 10.1016/s0020-7373(78)80049-2.
18. McCarthy, J. (1986). Applications of circumscription to formalizing common-sense knowledge. *Artificial intelligence*, 28(1), 89-116. DOI: 10.1016/0004-3702(86)90032-9.
19. McCarthy, J. (1978). History of LISP. *History of programming languages*, 173-185. DOI: 10.1145/800025.1198360.
20. Xue, H., Yang, Q. & Chen S. (2009). SVM: Support vector machines. The top ten algorithms in data mining. *Chapman and Hall/CRC*, 51-74. DOI: 10.1201/9781420089653-10.
21. Li, Z. et al. (2021). A survey of convolutional neural networks: analysis, applications, and prospects. *IEEE transactions on neural networks and learning systems*, 33(12), 6999-7019. DOI: DOI: 10.1109/TNNLS.2021.3084827.
22. Zhang, X. (2021). The AlexNet, LeNet-5 and VGG NET applied to CIFAR-10. *2nd International Conference on Big Data & Artificial Intelligence & Software Engineering (ICBASE). IEEE*, 414-419. DOI: 10.1109/ICBASE53849.2021.00083.
23. Feuerriegel, S. et al. (2024). Generative ai. *Business & Information Systems Engineering*, 66(1), 111-126. DOI: 10.2139/ssrn.4443189.
24. Creswell, A. et al. (2018). Generative adversarial networks: An overview. *IEEE signal processing magazine*, 35(1), 53-65. DOI: 10.1109/MSP.2017.2765202.

25. Mathew, A., Amudha, P. & Sivakumari, S. (2021). Deep learning techniques: an overview. *Advanced Machine Learning Technologies and Applications: Proceedings of AMLTA 2020*, 599-608. DOI: 10.1007/978-981-15-3383-9_54.
26. Ienca, M. (2023). On artificial intelligence and manipulation. *Topoi*, 42(3), 833-842. DOI: 10.1007/s11245-023-09940-3.
27. Yuste, R. (2015). From the neuron doctrine to neural networks. *Nature reviews neuroscience*, 16(8), 487-497. DOI: 10.1038/nrn3962.
28. Moerland, P. D. & Fiesler, E. (2020). Neural network adaptations to hardware implementations. *Handbook of neural computation, CRC Press*, E1. 2: 1-E1. 2: 13. DOI: 10.1887/0750303123/b365c78.
29. Kanal, L. N. (2003). Perceptron. *Encyclopedia of Computer Science*, 1383-1385.
30. Popescu, M. C. et al. (2009). Multilayer perceptron and neural networks. *WSEAS Transactions on Circuits and Systems*, 8(7), 579-588.
31. Chauhan, R., Ghanshala, K. K. & Joshi, R. C. (2018). Convolutional neural network (CNN) for image detection and recognition. *First international conference on secure cyber computing and communication (ICSCCC)*, 278-282. DOI: 10.1109/ICSCCC.2018.8703316.
32. Wu, J. (2017). Introduction to convolutional neural networks. *National Key Lab for Novel Software Technology. Nanjing University. China*, 5(23), 495.
33. Gu, J. et al. (2015). Recent advances in convolutional neural networks. *Pattern recognition*, 77, 354-377. DOI: 10.1016/j.patcog.2017.10.013.
34. Pinaya, W. H. L. et al. (2020). Convolutional neural networks. *Machine learning, Academic Press*, 173-191. DOI: 10.1016/B978-0-12-815739-8.00010-9.
35. Yu, Y. et al. (2019). A review of recurrent neural networks: LSTM cells and network architectures. *Neural computation*, 31(7), 1235-1270. DOI: 10.1162/neco_a_01199.
36. Mienye, I. D., Swart, T. G. & Obaido, G. (2024). Recurrent neural networks: A comprehensive review of architectures, variants, and applications. *Information*, 15(9), 517. DOI: 10.3390/info15090517.
37. Dhruv, P., Naskar, S. (2020). Image classification using convolutional neural network (CNN) and recurrent neural network (RNN): A review. *Machine learning and information processing: proceedings of ICMLIP 2019*, 367-381. DOI: 10.1007/978-981-15-1884-3_34.
38. Yinka-Banjo, C. & Ugot, O. A. (2020). A review of generative adversarial networks and its application in cybersecurity. *Artificial Intelligence Review*, 53, 1721-1736. DOI: 10.1007/s10462-019-09717-4.
39. Wang, K. et al. (2017). Generative adversarial networks: introduction and outlook. *IEEE/CAA Journal of Automatica Sinica*, 4(4), 588-598. DOI: 10.1109/JAS.2017.
40. Gui, J. et al. (2021). A review on generative adversarial networks: Algorithms, theory, and applications. *IEEE transactions on knowledge and data engineering*, 35(4), 3313-3332. DOI: 10.1109/TKDE.2021.3130191
41. Zhao, H. et al. (2021). Point transformer. *Proceedings of the IEEE/CVF international conference on computer vision*, 16259-16268. DOI: 10.1109/ICCV48922.2021.01595
42. Janiesch, C., Zschech, P. & Heinrich, K. (2021). Machine learning and deep learning. *Electronic markets*, 31(3), 685-695. DOI: 10.1007/s12525-021-00475-2
43. Shinde, P. P., Shah, S. (2018). A review of machine learning and deep learning applications. *Fourth international conference on computing communication control and automation (ICCUBEA), IEEE*. 1-6. DOI: 10.1109/ICCUBEA.2018.8697857.
44. Ongsulee, P. (2017). Artificial intelligence, machine learning and deep learning. *15th international conference on ICT and knowledge engineering (ICT&KE), IEEE*, 1-6.
45. Xu, F. et al. (2019). Explainable AI: A brief survey on history, research areas, approaches and challenges. *Natural language processing and Chinese computing: 8th cCF international conference, NLPCC, Springer International Publishing*, 563-574. DOI: 10.1007/978-3-030-32236-6_51.
46. Chaddad, A. et al. (2023). Survey of explainable AI techniques in healthcare. *Sensors*, 23(2), 634.
47. Gade K. et al. (2019). Explainable AI in industry. *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, 3203-3204. DOI: 10.1145/3292500.3332281.
48. Sanderson, K. (2023). GPT-4 is here: what scientists think. *Nature*, 615(7954), 773.
49. Da'u, A. & Salim, N. (2020). Recommendation system based on deep learning methods: a systematic review and new directions. *Artificial Intelligence Review*, 53(4), 2709-2748. DOI: 10.1007/s10462-019-09744-1.
50. Sielis, G. A., Tzanavari, A. & Papadopoulos, G. A. (2014). Recommender systems review of types, techniques, and applications. *Encyclopedia of Information Science and Technology, Third Edition, IGI Global Scientific Publishing*, 7260-7270.

51. Wirkuttis, N., Klein, H. (2017). Artificial intelligence in cybersecurity. *Cyber, Intelligence, and Security*, 1(1), 103-119.
52. Patil, P. (2016). Artificial intelligence in cybersecurity. *International journal of research in computer applications and robotics*, 4(5), 1-5.
53. Wiafe, I. et al. (2020). Artificial intelligence for cybersecurity: a systematic mapping of literature. *Ieee Access*, 8, 146598-146612. DOI: 10.1109/ACCESS.2020.3013145.
54. Radanliev, P. (2024). Artificial intelligence and quantum cryptography. Journal of Analytical Science and Technology, 15(1), 4. DOI: 10.1186/s40543-024-00416-6.
55. Chang, C. C. & Echizen, I. (2025). Steganography beyond space-time with chain of multimodal AI. *Scientific Reports*, 15(1), 12908.