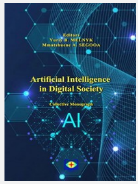
Check for updates

Artificial Intelligence in Digital Society,
Volume 1, 2026

DOI: 10.26697/9786177089192.2026
ISBN 978-617-7089-19-2 (Vol. 1)
ISBN 978-617-7089-18-5 (Series)



Chapter 1. The Evolution of the Theory and Practice of Artificial Intelligence

Stadnik A. V. [1,2], **Mykhaylyshyn U. B.** [1]

[1] *Uzhhorod National University, Ukraine*
[2] *Kharkiv Regional Public Organization "Culture of Health", Ukraine*

Received: *01.12.2025;* ***Accepted:*** *10.02.2026;* ***Published:*** *10.03.2026*

Abstract

The authors first review the historical milestones in the emergence and development of artificial intelligence (AI). They demonstrate that progress is rarely linear, and today's successes do not guarantee solutions to future challenges. Further research shows that the theoretical foundation of AI is a multi-layered, interdisciplinary structure that includes biological, cognitive, philosophical, mathematical, and logical aspects, as well as machine learning, neural networks, and natural language processing. Modern artificial intelligence is the result of the interaction of these theoretical branches, each of which contributes to the effective operation of intelligent systems. The authors conclude that the evolution of AI from classical symbolic to machine learning represents a fundamental shift in the understanding and construction of artificial intelligence. However, a consideration of symbolic and statistical approaches shows that neither is ideal, often requiring hybrid solutions that combine multiple methods.

Keywords: evolution of artificial intelligence, machine learning, symbolic artificial intelligence, neural networks, machine learning, intelligent systems.

Cite this chapter as:
Stadnik, A. V., & Mykhaylyshyn, U. B. (2026). The evolution of the theory and practice of artificial intelligence. In Y. B. Melnyk & M. A. Segooa (Eds.), *Artificial Intelligence in Digital Society, Vol. 1* (pp. 11–26). KRPOCH. https://doi.org/10.26697/aids.2026.1
The electronic version of this chapter is complete. It can be found online in the AIDS Archive https://doi.org/10.26697/aids

Historical Milestones in Artificial Intelligence Research

Artificial Intelligence (AI) is a system that enables machines (computers, neural networks, etc.) to solve problems similar to those of the human mind, including perception, learning, reasoning, and decision-making, by analogy with human cognitive functions (Gacem & Aouane, 2024).

The preconditions for AI were formed long before the appearance of the first electronic computers. The possibility of creating and developing artificial humans, higher intelligence, and superintelligence capable of thinking better than humans has been a topic of discussion for quite some time. Ancient Greek myths about Hephaestus (god of fire and forge) described automatons (animate, metal statues of animals, people, and monsters). For example, Talos was a giant sculpted from bronze by Hephaestus who patrolled the island of Crete, protecting it from pirates; the Halkotaurus were bronze fire-breathing bulls (Fan et al., 2020). In medieval Islamic science, the Banu Musa brothers and Al-Jazari developed various automatic devices, including water clocks, self-playing musical mechanisms, and servomechanisms. Particularly revealing are Al-Jazari's "android" figures, which essentially represented early models of programmable automata containing feedback elements (Hill, 1991).

The concept of modern scientists was reduced to a partial imitation of the computational capabilities of the human brain through mathematical models, which subsequently led to the invention of mechanical computing machines:

1623 – Wilhelm Schickard came up with the "Counting clock" – the first adding machine capable of performing four arithmetic operations. The mechanism's operation was based on the use of stars and gears.

1642 – Blaise Pascal created "Pascal's calculator", which could only perform addition and subtraction.

1673 – Gottfried Leibniz's arithmometer, a mechanical calculating machine capable of addition, subtraction, multiplication, and division (Chang, 2020).

The development of autonomous mechanical computing devices became the prototype of modern AI technology and depended on the existing technological and computational capabilities of a particular period.

In the first half of the 20th century, numerous publications appeared devoted to the theoretical prerequisites for the creation of AI:

1920 – Czech playwright Karel Čapek released a science fiction play, "RUR", in which he proposed the idea of artificial humans, which he called robots (Chang, 2020).

1943 – McCulloch & Pitts published "A Logical Calculus of the Ideas Inherent in Nervous Activity", describing the first mathematical model of a neural network (McCulloch & Pitts, 1943).

1948 – Claude Shannon, in his work "Mathematical theory of communication", proposed to consider information as something new that is transmitted during communication. Information is measured in "binary digits" (0 or 1), better known as "bits" (Shannon, 1950).

In the second half of the 20th century, the term "artificial intelligence" was coined and entered into common usage, and interest in AI reached its peak:

1950 – Alan Turing published the book "Computing Machinery and Intelligence". He proposed the "Turing test", which asks whether a machine can fool a scientist into thinking it is communicating with a human (Turing, 1950).

1952 – Arthur Samuel wrote a computer checkers program that played at a master level with the ability to learn as you play (Samuel, 2000).

1956 – John McCarthy, Marvin Minsky, Nathaniel Rochester & Claude Shannon organized the Dartmouth Workshop on Artificial Intelligence, where the term was first used. Participants suggested that machine learning, language use, and problem-solving were problems that could be solved in the near future (McCarthy et al., 1955).

Then it continued, the creative development of artificial intelligence, from programming languages still in use today to books and films exploring the idea of robots:

1958 – John McCarthy created LISP (LISt Processing), the first programming language for AI research, which is still popular today (Moor, 2006).

1959 – Arthur Samuel coined the term "machine learning" (Samuel, 2000).

1961 – The first industrial robot, Unimate, began working on the assembly line at a General Motors plant in New Jersey.

1965 – Edward Feigenbaum & Joshua Lederberg created the first "expert system", which was a form of AI programmed to replicate the thinking and decision-making abilities of human experts.

1966 – Joseph Weizenbaum created the first "chatbot" ELIZA, a pseudo-psychotherapist that used Natural Language Processing (NLP) for communicating with people.

1979 – The Association for the Advancement of Artificial Intelligence (AAAI) was founded (AAAI, 2025).

Next, a period of rapid growth and decline of interest in AI has begun due to the lack of technological breakthroughs and reduced funding:

1980 – The first AAAI conference was held at Stanford (AAAI, 2025).

1980 – The first expert system, known as XCON (Expert Configurator), appeared on the commercial market. It was designed to assist in ordering computer systems by automatically selecting components according to the customer's needs.

1985 – An autonomous drawing program known as AARON is demonstrated at the AAAI conference.

1986 – Ernst Dieckmann and his team from the Bundeswehr University of Munich created and demonstrated the first driverless car (robomobile). It could reach speeds of up to 80 km/h on roads without obstacles or human drivers.

1987 – Alacrity was commercially launched by Alacrity Inc. Alacrity was the first strategic management consulting system and used a sophisticated expert system.

1987 – The market for specialized LISP-based hardware collapsed due to the emergence of cheaper and more accessible competitors capable of running LISP software, including products from IBM and Apple.

1988 – Rollo Carpenter invented the chatbot Jabberwacky, which he programmed to carry on interesting and engaging conversations with people (AAAI, 2025).

Despite the lack of funding, the early1990s saw impressive advances in AI research, including the creation of the first AI system capable of defeating a reigning world chess champion, the introduction of AI into everyday life (the first Roomba robot vacuum cleaner, the first commercial speech recognition software on Windows computers):

1997 – Deep Blue (IBM) defeated world chess champion Garry Kasparov, becoming the first program to beat a human world chess champion;

1997 – Windows released speech recognition software (Dragon Systems);

2000 – Professor Cynthia Breazeale has developed the first robot capable of imitating human emotions using a face including eyes, eyebrows, ears, and mouth;

2002 – The first Roomba robot vacuum cleaner was released;

2003 – NASA landed two rovers (Spirit & Opportunity) on Mars, and they explored the planet's surface without human intervention;

2006 – Twitter, Facebook, and Netflix began using AI as part of their advertising and user experience algorithms;

2010 – Microsoft released the Xbox 360 Kinect, the first gaming device designed to track body movements and convert them into gaming commands;

2011 – Apple released Siri, the first popular virtual assistant (AAAI, 2025; Chang, 2020).

At this time, digital product is becoming one of the most rapidly developing areas (Pypenko, 2019). We are witnessing a surge in the popularity of AI tools, including virtual assistants and search engines. Deep learning and big data have also gained traction during this period:

2012 – Jeff Dean and Andrew Ng trained a neural network to recognize cats;

2016 – Hanson Robotics has created a humanoid robot named Sofia with a realistic human appearance and the ability to see and reproduce emotions, as well as communicate;

2017 – Facebook programmed two AI-powered chatbots to communicate and learn how to negotiate, but during the process of communicating, they eventually abandoned English and began developing their own language, completely autonomously (Chang, 2020);

2019 – Google's AlphaStar reached Grandmaster level in the video game StarCraft 2, outperforming almost all human players;

2020 – OpenAI began beta testing GPT-3, a deep learning model. It became the first model to produce content virtually indistinguishable from human-created content;

2021 – OpenAI developed DALL-E, which can process and understand images at a level sufficient to generate accurate captions (AAAI, 2025; Chang, 2020);

2023 – In the Academic International Corporation, the problems of legitimising AI-based ChatBots in scientific research were studied, and an attribution (AIC AI Chatbots) was developed, which is proposed for use in indicating the role and level of involvement of AI and ChatBots in research and publications (Melnyk & Pypenko, 2023);

2025 – Large Language Models Pass the Turing Test (ELIZA, GPT-4o, LLaMa-3.1-405B, and GPT-4.5) (Jones & Bergen, 2025).

Thus, while modern achievements are impressive, they also remind us of the need for a sober assessment of the capabilities and limitations of AI technology. History teaches us that progress is rarely linear, and today's successes do not guarantee solutions to tomorrow's challenges. Fundamental questions about the nature of intelligence and its understanding remain open. The future of AI will be determined not only by technical breakthroughs but also by society's ability to develop ethical frameworks, regulatory mechanisms, and social institutions for the responsible development and application of this technology.

Theoretical Foundations of Artificial Intelligence Systems

The theoretical aspects of AI systems are diverse and encompass *biological, cognitive, philosophical, mathematical, logical, machine learning, neural networks, and natural language processing*. They reflect different views on intelligence, cognition, and system operations.

Biological Aspects. The origins of AI, as well as its history, are closely linked to brain sciences (neurophysiology, anatomy and physiology of the nervous system, psychology, etc.). Many of the founding scientists of AI are also brain scientists, and many discoveries in AI are based on biological research. For example, working memory, which was discovered using magnetic resonance imaging, inspired the development of the memory module in machine learning models, ultimately leading to the creation of the long short-term memory (LSTM) model. Changes in the spinal cord that occur during learning have inspired the creation of a new algorithm, *Elastic Weight Consolidation* (EWC), for continuous learning. Neural connections in the human brain, discovered using a microscope, inspired the development of artificial neural networks (Fan et al., 2020).

The goal of AI is to develop computer systems capable of performing tasks traditionally performed by human intelligence, with functions such as information reception, processing, decision-making, and control. The goal of brain science is to study the structure, functions, and operating mechanisms of the biological brain (perception, recognition of multisensory information, and decision-making regarding interactions with the environment). Thus, AI is very similar to human intelligence and can be considered a simulation of the human brain's cognitive abilities (Miśkiewicz, 2019). A comparison of human intelligence and AI has revealed that, although both systems are capable of performing similar tasks, their underlying mechanisms and limitations are fundamentally distinct (Table 1).

Table 1.1

A Final Comparison of Human Intelligence and Artificial Intelligence

Characteristic	Human Intelligence	Artificial Intelligence
Warp	Biological (neural activity, neuromodulation, and metabolic restrictions)	Digital and machine (logical operations, energy-efficient matrix multiplication, and discrete data representation)
Information processing	Analog and discrete	Discrete
Types of errors	Biologically stochastic (random)	Systemic (mathematically determined)
Consciousness	Possesses consciousness or self-awareness	Does not possess consciousness or self-awareness
Thinking	Abstract and heuristic based on experience, emotion, and cultural context	Heuristic based on computations, search algorithms, probabilistic inferences
Adaptation and decision making	Based on life experience, emotions, and cultural context	Based on pre-programmed algorithms and data, they often require reprogramming
Creation	Based on associative thinking, life experience, emotional, and cultural context	Based on patterns and the recombination of information from training data
Embodiment and Cognition	They develop together, sensorimotor feedback shapes learning, and bodily experience influences abstract thinking.	When integrated into robotic systems, embodiment is not phenomenologically or biologically equivalent.
Education	Possibly one-shot learning and generalization based on minimal information	Repetitive learning from new large data sets

Thus, AI and human intelligence share standard features in the ability to learn and adapt, but differ significantly in their nature, operating mechanisms, and capabilities.

Cognitive Psychology helps understand and model thought and perception processes. The mental aspects of AI are rooted in an attempt to describe human thinking as an information-processing process. Digital AI models are partly inspired by theories of cognitive science, which study perception and processing of sensory information, memory, attention, logical and associative thinking, and decision-making.

Fundamental components of AI cognitive constructs are knowledge representation mechanisms. Several approaches have emerged in artificial intelligence: *the symbolic approach* (knowledge is represented in the form of logical structures and rules); *the subsymbolic approach* (knowledge is stored in distributed representations characteristic of neural networks); and *the hybrid approach* (a combination of logical structures and neural network learning). Each approach reflects specific views on the functioning of the brain's cognitive systems, including the processing of contexts, connections between objects, and the ability to generalize.

Modern AI systems demonstrate functional analogs of human cognitive processes: Perception is realized through computer vision models, speech recognition, and multimodal transformers. Attention is mathematically formalized in the self-attention mechanism (self-attention), which allows us to highlight the most significant elements of the input data. Memory is represented by internal state structures (LSTM, GRU), external storage (Differentiable Neural Computers), or a quasi-semantic, multidimensional vector space (embedding space). Although such systems do not reproduce the biological mechanisms of intelligence, they perform similar functions, which allows them to be considered as cognitive models (Achler, 2024).

The Philosophy of AI examines the nature, consciousness, and ethical implications of utilizing AI. For example, it asks questions such as, Can a machine think like a human? Does it possess consciousness and subjective experience? A modern view of the Turing Test, according to the physical-symbolic systems hypothesis (Newell & Simon, 1971), argues that symbol manipulation is sufficient for intelligence. Meanwhile, John Searle's "Chinese Room" argument demonstrates that a program manipulates symbols syntactically but does not comprehend their semantics, much like a person in a room following instructions in Chinese without understanding the language. This contradicts the idea that the brain works like a computer (Searle, 1993).

Philosophical aspects of AI also include issues of consciousness, free will, the alignment of human values, ethics, and bias. Key issues include whether a machine can possess consciousness, how to ensure ethical behavior in AI, and how to integrate human values into algorithms without distortion or bias. Furthermore, the problem of explaining AI decisions and controlling its actions in society is discussed. To address these issues, an interdisciplinary approach that brings together philosophers, scientists, and engineers is essential, as is the development of transparent and interpretable algorithms (Gacem & Aouane, 2024).

Mathematical Aspects of AI encompass *probability theory, mathematical statistics, linear algebra, and optimization algorithms*.

Linear algebra provides tools for representing and processing data using vectors and matrices. These structures play a key role in the development of machine learning models and neural networks.

Vectors and matrices are fundamental elements of representing data in feature space. Eigenvalues and eigenvectors play a crucial role in dimensionality reduction and data analysis using methods such as principal component analysis (PCA) (Jolliffe, 2002). These concepts enable the extraction of the most relevant information from large and complex datasets.

Probability theory enables us to model the uncertainty and randomness inherent in data. *Statistical methods* are essential for evaluating models and drawing conclusions from data (Casella & Berger, 2002). Probability distributions describe the distribution of values of a random variable, which is essential when modeling stochastic processes. Bayesian theory and probabilistic models enable us to update probabilities based on new data. This is particularly useful when developing adaptive systems that learn in dynamic environments.

Optimization algorithms are at the core of the training process for AI models. Gradient descent is used to minimize the loss function by updating parameters in the direction of the steepest slope (Ruder, 2016). Stochastic gradient descent (SGD) is a variation of the method that uses random subsets of the data to speed up training, which is particularly useful for large datasets (Almudevar, 2021).

Logic*,* along with mathematical methods, plays a crucial role in representing knowledge, formalizing reasoning, and constructing logical programming systems. These aspects enable the creation of inference, deduction, and proof algorithms used by many AI systems. Logical approaches are based on various types of formal logic, including classical propositional logic, predicate logic, and specialized description logics. These methods allow knowledge and rules to be precisely described in the form of formal statements. Knowledge bases and production systems are used to represent knowledge, and the reasoning process is implemented using inference algorithms, such as resolution. The logical programming language Prolog serves as a practical example of the application of logical foundations in AI, where rules and facts are specified declaratively, and inference is performed automatically (Genereth & Nilsson, 1987).

A key element of the logical foundations of AI is inference algorithms, which enable the generation of new knowledge from a given set of facts and rules. The main goal is not simply to obtain an answer, but to explain its logic, preserving the reasoning tree and the possibility of counterfactual analysis. Such approaches form the basis of explainable AI (XAI), which is becoming increasingly popular in modern systems. This ensures the transparency and trust in intelligent systems (Genereth & Nilsson, 1987).

In recent years, there has been an increasing integration of logical and algorithmic methods with machine learning. Hybrid neural-symbolic architectures combine the ability of neural networks to process unstructured data with the ability of logic to provide formal reasoning and explanation. Such systems use logic to ask "why" and the conditions under which decisions are made, while neural networks address the question of "what to predict". These approaches open up new possibilities for creating more adaptive, simultaneously explainable, and formally verifiable intelligent statements.

Machine Learning is a core component of artificial intelligence, enabling systems to learn from data, make predictions, or make decisions without being explicitly programmed (Mitchell, 1997). There are three primary learning paradigms: *supervised, unsupervised, and reinforcement learning*.

In *supervised learning*, models are trained on labeled data, where each input is associated with a corresponding output value. Regression predicts continuous values, while classification categorizes data into discrete categories using methods such as logistic regression and support vector machines (SVM) (Bishop, 2006).

Unsupervised learning methods work with unlabeled data and aim to discover hidden structures or patterns within it. Clustering groups similar data, for example, using the k-means method. Dimensionality reduction methods, such as PCA, reduce data complexity while preserving important information (Hastie, Tibshirani, & Friedman, 2009).

Reinforcement learning involves training an agent by interacting with its environment and learning from the consequences of its actions, which are mediated by rewards or punishments (Sutton & Barto, 2018). This approach is efficient in problems where the sequence of decisions influences the final goal.

Neural Networks are models inspired by biological neurons that can learn complex functions and representations (Goodfellow, Bengio, & Courville, 2016). Artificial neurons are basic devices that take input and generate output through an activation function. Multilayer perceptrons (MLPs) are neural networks with one or more hidden layers that can model nonlinear relationships. The universal approximation theorem states that neural networks with sufficient neurons can approximate any continuous function (Hornik, Stinchcombe, & White, 1989). Convolutional neural networks (CNNs) specialize in image processing by capturing spatial dependencies (LeCun et al., 2015). Recurrent neural networks (RNNs) are well-suited for sequential data, such as text and audio, because they account for temporal dependencies.

Natural Language Processing *(NLP)* is another crucial AI task enabled by computational linguistics. The theoretical aspects of NLP encompass *syntactic,* *semantic, and pragmatic analysis, statistical text processing methods, and consideration of linguistic constructs*. This enables the creation of intelligent translation systems, conversational interfaces, and assistants that understand and generate spoken language.

NLP process involves several stages.

1. Data entry involves receiving text or voice data.

2. Pre-processing involves cleaning and structuring data (e.g., tokenization, stop word removal).

3. Meaning extraction involves using machine learning algorithms to analyze the meaning of words in context, disambiguate, and infer user intent (Alammar, 2025).

4. An appropriate response or performing a task based on the extracted value (Manning et al., 2020).

Modern NLP is based on highly efficient computational models: word embedding learning: word2vec, GloVe – represent words as vectors encoding semantic similarity; ELMo, BERT - take into account the dependence of word meaning on context; *transformers* - provide parallel processing of sequences and deep contextual understanding; *large language models* (LLM) – trained on huge data and capable of performing a variety of tasks without specialized tuning (Ethayarajh, 2019).

Thus, the theoretical foundations of AI form a multilayered, interdisciplinary structure, incorporating biological, cognitive, philosophical, mathematical, and logical aspects, as well as machine learning, neural networks, and natural language processing. Modern AI is the result of the interaction of these theoretical fields, each contributing to our understanding of the nature of intelligence and the construction of effective intelligent systems.

From Symbolic Artificial Intelligence to Machine Learning: Paradigm Shifts

Problems in AI can be solved in different ways. Some approaches (symbolic) assume that the mind is a system of symbolic representations and logical manipulations of them; others (statistical and machine learning) interpret intelligence as the ability to extract patterns from data and optimize decisions. The transition from the first to the second approach was not immediate. It represented a gradual shift in emphasis, an interweaving of methods, and a reflection on which properties of intelligence are more important for practice.

Symbolic AI (Good Old-Fashioned AI (GOFAI) is based on the Newell and Simon hypothesis (1971), which posits that intelligence can be modeled through the manipulation of symbols that express knowledge and the logical relationships between them. The methodology is based on constructing productive rules ("if-then") that connect symbols into logical relationships. Using these rules, systems draw conclusions, form hypotheses, and determine what additional information to request. This structure enables the modeling of cognitive processes with consistent logic and adaptation based on knowledge, rather than relying solely on statistics.

This methodology also draws on biological analogies, where the model includes neurons (perceiving and discriminating) that are hierarchically organized into structures capable of complex inference. Symbolic AI creates models that closely resemble the cognitive processes of living organisms, facilitating the explanation of results and the analysis of critical factors influencing decisions. Furthermore, the reasoning process is transparent and traceable, facilitating its understanding and debugging. Rule-based systems are less demanding on computing resources. Symbolic AI can often run on computer CPUs, making it more energy-efficient than data-intensive machine learning approaches that typically require powerful GPUs.

The main tools of symbolic AI are:

- knowledge bases containing a set of rules and facts in the form of symbols;
- means of formalizing knowledge (for example, production rules; logical expressions, semantic networks, frames;
- logical interpreters and inference engines that apply rules to reason and make decisions;
- formalized knowledge representation languages and explanation systems that provide logic traceability (Prolog and other logical languages) (Garrido-Merch & Puente, 2025).

Symbolic AI, dominant from the 1950s to the 1980s, used hand-crafted rules and logical reasoning to model intelligence. It was effective for tasks requiring explicit knowledge representation, such as expert systems, but struggled with the ambiguity, scalability, and complexity of the real world.

The 1980s and 1990s saw the modernization and growth of computing power, enabling greater use of large databases and statistical methods that allow machines to learn based on patterns in the data they receive, rather than relying solely on predefined rules. This made it possible to process probabilistic, unstructured, and variable data, which was difficult for symbolic methods.

This paved the way for a radical paradigm shift: the focus shifted from explicit, transparent, and precise knowledge to data and optimization. This led to the emergence and rise of machine learning (Goodfellow, Bengio & Courville, 2016).

Machine Learning is a set of methods where models extract patterns from examples. Machine learning is based on *statistics, information theory, and computational learning theory*. The key idea is that instead of explicitly programming rules for solving a problem, a system should automatically extract patterns from data. This represents a shift in emphasis from algorithm design to the design of learning architectures and the collection of relevant data (Bishop, 2006).

The primary methods of machine learning are *defining the model architecture, loss function, optimization algorithm, and validation procedure*. Machine learning's key strengths include adaptability and scalability, robustness to noise and partial data distortion, and practical efficiency, often exceeding human performance in applied tasks.

The most dramatic paradigm shift has occurred in the last fifteen years, with the resurgence and triumph *of deep neural networks*. Although the basic algorithms had existed for decades, the convergence of several factors led to a qualitative leap in their performance. The availability of massive datasets, thanks to the internet and digitalization, provided training material. The computing power of graphics processing units made it possible to train networks with billions of parameters. Algorithmic innovations, such as batch normalization, attention mechanisms, and residual connections, have addressed the challenges of training deep networks. Breakthrough results followed one after another. In 2012, the AlexNet convolutional neural network radically outperformed traditional methods in the ImageNet image recognition competition. Machine translation systems based on recurrent networks and attention mechanisms achieved near-human quality. Image and text generation models demonstrated impressive creative capabilities (Krizhevsky et al., 2012).

The emergence of the Transformer architecture in 2017, along with the subsequent development of large-scale language models, was particularly significant. Systems (GPT, BERT, and others) have demonstrated that pre-training on massive text data, followed by fine-tuning, can create models with surprisingly broad natural language processing capabilities, without requiring explicit encoding of linguistic rules (Ethayarajh, 2019).

The shift from symbolic AI to machine learning reflects deep, fundamental disagreements about the nature of knowledge and intelligence. The symbolic approach assumes that intelligence requires explicit, declarative representations of knowledge and its manipulation through logical rules. Machine learning, intense learning, embodies the position that knowledge emerges from experience and does not necessarily have an explicit symbolic form. Expertise can exist without explicit representation of rules; knowledge can be embodied and procedural (Gorner, 2007).

Methodologically, the two approaches differ in their method of problem-solving. *Symbolic AI* follows a top-down strategy: problem analysis, decomposition into subtasks, knowledge formalization, and construction of reasoning algorithms. This is an engineering approach, where each system component is designed to perform a specific function. *Machine learning* is primarily a bottom-up approach, involving the selection of an appropriate architecture, data collection, model training, and model validation on independent data. The inner workings of a trained model often remain opaque; knowledge is extracted inductively from statistical regularities rather than deductively constructed.

Symbolic AI retains its advantages in tasks that require explicit reasoning, the explanation of decisions, guaranteed correctness, and working with a small number of examples. Symbolic systems naturally support composition, systematic thinking, and the integration of heterogeneous knowledge sources. Their behavior is predictable and verifiable. *Machine learning* excels in tasks of pattern recognition, working with noisy data, generalization to new situations, and processing sensory information. It naturally scales to big data and automatically adapts to changes in the data distribution. Formalization of costly expert knowledge is not required.

However, machine learning also has significant weaknesses. Models can be opaque black boxes, making them difficult to understand and debug. They require large volumes of labeled data and computational resources. Generalization is limited by proximity to the training distribution, and models are vulnerable to adversarial examples. Integrating symbolic knowledge and common sense remains a challenge.

The big data revolution has fundamentally changed the AI landscape. The exponential growth of computing power, particularly the advent of specialized accelerators like GPUs, has enabled the training of models with billions of parameters. What seemed computationally infeasible just twenty years ago is now routinely performed in research labs and even on personal computers.

The commercial success of machine learning has attracted massive investment from tech companies and venture capital funds, creating a powerful economic incentive for further research and development. The availability of open data, code, and pre-trained models has made a positive feedback loop, accelerating progress. The integration of machine learning into educational programs and the development of specialized training courses have led to an influx of talent into the field, further accelerating its growth (Sifatkaur et al., 2023).

However, due to the lack of explainability and opacity of purely statistical models, the field of neuro-symbolic AI, which combines the advantages of both approaches, has recently developed. This hybrid approach seeks to combine the adaptability and ability to learn from machine learning data with the logical rigor and explainability of symbolic systems, which is particularly important for applications in mission-critical areas (Gacem & Aouane, 2024).

Thus, the evolution of AI from classical symbolic to machine learning represents a fundamental shift in the understanding and construction of artificial intelligence – from manual knowledge processing to systems capable of autonomous learning and adaptation, while maintaining the desire for transparency and controllability of decisions. The transition from symbolic to statistical approaches shows that neither provides all the answers. First, symbolic systems have proven insufficient for understanding the full complexity of human language.

Statistical systems have succeeded not by replacing symbolic representations, but by finding new ways to describe patterns that symbolic approaches struggled to formalize. Second, practical AI problems often require hybrid solutions based on multiple approaches. Third, the qualities of symbolic systems (interpretability and explainability) sometimes outweigh all the advantages of modern neural systems.

References

Achler, T. (2024). What AI, neuroscience, and cognitive science can learn from each other: An embedded perspective. *Cognitive Computation, 16,* 2428–2436. https://doi.org/10.1007/s12559-023-10194-9

Alammar, J. (2025). Large language models: Architecture and training. From next-word prediction to reasoning. *In Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.*2 (KDD '25). Association for Computing Machinery, New York. USA. https://doi.org/10.1145/3711896.3736805

Almudevar, A. (2021). *Theory of statistical inference* (1st ed.). Chapman and Hall/CRC. https://doi.org/10.1201/9781003049340

Association for the Advancement of Artificial Intelligence. (2025). *About the Association for the Advancement of Artificial Intelligence (AAAI).* https://aaai.org

Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer, Berlin. https://link.springer.com/book/9780387310732

Casella, G., & Berger, R. L. (2002). *Statistical inference* (2nd ed.). Duxbury Press, Pacific Grove. https://pages.stat.wisc.edu/~shao/stat610/Casella_Berger_Statistical_Inference.pdf

Chang, A. C. (2020). History of artificial intelligence. In *Intelligence-Based Medicine, Artificial Intelligence and Human Cognition in Clinical Medicine and Healthcare* (pp. 23–27). https://doi.org/10.1016/B978-0-12-823337-5.00002-0

Ethayarajh, K. (2019). *How contextual are contextualized word representations? Comparing the geometry of BERT, ELMo, and GPT-2 embeddings.* ArXiv. https://doi.org/10.48550/arXiv.1909.00512

Fan, J., Fang, L., Wu, J., Guo, Y., & Dai, Q. (2020). From brain science to artificial intelligence. *Engineering, 6*(3), 248–252 https://doi.org/10.1016/j.eng.2019.11.012

Gacem, H., & Aouane, A. (2024). Conceptual foundations of artificial intelligence. *Journal of El-Manhel Economy*, *7*(1), 1215–1224. https://asjp.cerist.dz/en/article/247713

Garrido-Merchán, E., & Puente, C. (2025). *GOFAI meets Generative AI: Development of expert systems by means of large language models.* ArXiv. https://doi.org/10.48550/arXiv.2507.13550

Genesereth, M. R., & Nilsson, N. J. (1987). *Logical foundations of artificial intelligence*. Los Altos, CA: Morgan Kaufmann. https://kryten.mm.rpi.edu/COURSES/LOGAIS02/nonmonotonic.pdf

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning.* MIT Press. https://www.deeplearningbook.org/

Gorner, P. (2007). Authenticity. In *Heidegger's Being and Time: An Introduction* (pp. 105–152). Cambridge University Press. https://doi.org/10.1017/CBO9780511808036.009

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The element of statistical learning: data mining, inference, and prediction.* Springer, Berlin. https://doi.org/10.1007/978-0-387-84858-7

Hill, D. (1991). Mechanical engineering in the Medieval Near East. *Scientific American*, *264*(5), 100–105. https://www.jstor.org/stable/24936907

Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, *2*(5)*,* 359–366. https://doi.org/10.1016/0893-6080(89)90020-8

Jolliffe, I. T. (2002). *Principal component analysis* (2nd ed.). New York: Springer-Verlag. https://link.springer.com/book/10.1007/b98835

Jones, C. R., & Bergen, B. K. (2025). *Large language models pass the Turing test.* ArXiv. https://doi.org/10.48550/arXiv.2503.23674

Krizhevsky, A., Sutskever, I., & Hinton, G. (2012). ImageNet classification with deep convolutional neural networks. In F. Pereira, C. J. Burges, L. Bottou, & K. Q. Weinberger (Eds.), *Advances in Neural Information Processing Systems 25 Proceedings* (pp. 1097–1105). Curran Associates Inc. https://papers.nips.cc/paper_files/paper/2012

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521,* 436–444. http://dx.doi.org/10.1038/nature14539

Manning, C., Clark, K., Hewitt, J., Khandelwal, U., & Levy, O. (2020). Emergent linguistic structure in artificial neural networks trained by self-supervision. *Proceedings of the National Academy of Sciences, 117*(48), 30046-30054. https://doi.org/10.1073/pnas.1907367117

McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (1955). *A proposal for the Dartmouth summer research project on artificial intelligence.* http://jmc.stanford.edu/articles/dartmouth/dartmouth.pdf

McCulloch, W., & Pitts, W. (1943). A logical calculus of ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics, 5,* 115–133. https://doi.org/10.1007/BF02478259

Melnyk, Yu. B., & Pypenko, I. S. (2023). The legitimacy of artificial intelligence and the role of ChatBots in scientific publications. *International Journal of Science Annals, 6*(1), 5-10. https://doi.org/10.26697/ijsa.2023.1.1

Miśkiewicz, J. (2019). The merger of natural intelligence with artificial intelligence, with a focus on Neuralink company. *Virtual Economics*, *2*(3), 22–29. https://doi.org/10.34021/ve.2019.02.03(2)

Mitchell, T. (1997). *Machine learning*. McGraw-Hill Higher Education, New York. https://www.cs.cmu.edu/~tom/files/MachineLearningTomMitchell.pdf

Moor, J. (2006). The Dartmouth college artificial intelligence conference: The next fifty years. *AI Magazine*, *27*(4), 87–89. https://doi.org/10.1609/aimag.v27i4.1911

Pypenko, I. S. (2019). Digital product: The essence of the concept and scopes. *International Journal of Education and Science, 2*(4), 56. https://doi.org/10.26697/ijes.2019.4.41

Samuel, A. L. (2000). Some studies in machine learning using the game of checkers. *IBM Journal of Research and Development*, *44*(1.2), 206–226. https://doi.org/10.1147/rd.441.0206

Searle, J. R. (1993). Consciousness, explanatory inversion, and cognitive science. *Behavioral and Brain Sciences*, *13*(4), 585–596. https://doi.org/10.1017/S0140525X00080304

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*(3), 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Sifatkaur, D., Manmeet, S., Vaisakh S. B., Neetiraj, M., & Sukhpal, S. (2023). Mind meets machine: Unravelling GPT-4's cognitive psychology. *Bench Council Transactions on Benchmarks, Standards and Evaluations, 3*(3), Article 100139. https://doi.org/10.1016/j.tbench.2023.100139

Simon, H. A., & Newell, A. (1971). Human problem solving: The state of the theory in 1970. *American Psychologist, 26*(2), 145–159. https://doi.org/10.1037/h0030806

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction* (2nd ed.). A Bradford Book, the MIT Press. https://web.stanford.edu/class/psych209/Readings/SuttonBartoIPRLBook2ndEd.pdf

Turing, A. M. (1950). Computing machinery and intelligence. *Mind, LIX*(236), 433–460. https://doi.org/10.1093/mind/LIX.236.433

Information about the authors:

Stadnik Anatoliy Volodymyrovych – https://orcid.org/0000-0002-1472-4224; Doctor of Philosophy in Medicine, MD, Affiliated Associate Professor, Kharkiv Regional Public Organization "Culture of Health", Kharkiv, Ukraine; Uzhhorod National University, Uzhhorod, Ukraine.

Mykhaylyshyn Ulyana Bohdanivna – https://orcid.org/0000-0002-0225-8115; Doctor of Psychological Sciences, Full Professor; Head of the Department of Psychology, Uzhhorod National University, Uzhhorod, Ukraine.