UDC 004. 032.26:007:53:165.8 **doi:** https://doi.org/10.20998/2522-9052.2025.4.13

Yurii Parzhyn[1], Mykyta Lapin[2], Kostiantyn Bokhan[2]

[1] Augusta University, Augusta, USA
[2] National Technical University "Kharkiv Polytechnic Institute", Kharkiv, Ukraine

A NEW APPROACH TO BUILDING ENERGY MODELS OF NEURAL NETWORKS

Abstract. **Relevance.** Modern artificial neural network models require significant energy and other resources for training and operation. Training generative models involves vast amounts of data. At the same time, these models face challenges related to the trustworthiness of the information they generate. An alternative to current paradigms of building and training neural networks is the development of energy-based models, which could potentially overcome these shortcomings and bring information processing closer to biologically and physically grounded processes. However, existing energy-based models differ little from classical models in terms of their limitations and drawbacks. Therefore, developing new approaches to modeling energy-based information processing in neural networks is highly relevant. **The object of research** is the process of information processing in artificial neural networks. **The subject of the research** is the mathematical models for the construction and training of artificial neural networks. **The purpose of this paper** is to develop and experimentally validate a theoretical framework that postulates the energetic nature of information and its role in the self-organization and evolution of complex information systems. **Research Results.** A fundamental theory is proposed, describing information as a structure of perceived external energy parameters that govern the processes of forming the internal energetic structure of a system—its model of the external world. This theory encompasses concepts of energy landscapes, principles of energy-based structural and parametric reduction, and a critical analysis of existing computational paradigms. Experimental studies on the construction and training of the developed energy-based model confirm its high generalization ability in one-pass training without using the backpropagation algorithm on ultra-small training datasets.

Keywords: energy-aware computing; energy landscape; artificial neural networks; neurocomputing; information; entropy.

Introduction

Recent advances in the field of artificial intelligence, particularly in deep learning and artificial neural networks (ANNs), have led to significant progress in solving complex problems. However, the widespread application of these technologies has also revealed a number of fundamental limitations and shortcomings. These raise questions about the feasibility of creating truly autonomous, adaptive, and intelligent systems capable of self-organization and flexible interaction with a complex, dynamic environment, in a manner similar to biological organisms.

Examples of such limitations include: the necessity for massive amounts of data, along with substantial time, computational, and human resources required for training and operating complex models [1–3]; considerable energy consumption and the use of potable water for device cooling, resulting in environmental damage [4, 5]; and the frequent occurrence of errors and "hallucinations" in generative models, which significantly reduce trust in their outcomes, necessitate additional validation efforts, and severely limit their practical applications [6, 7].

We believe that these limitations stem from inherent constraints within the existing conceptual framework, which are not temporary but fundamental in nature. This assumption aligns with the views of several prominent researchers, including Yann LeCun, Gary Marcus, Anthony M. Zador, and others [8-10]. The findings of these and many other studies indicate that the primary fundamental limitations of modern ANNs are rooted in their statistical nature of learning and the rigid architecture of the systems. These are weakly connected to the deep physical principles of information processing and world modeling. In particular, these limitations lead to the fundamental phenomenon of the “entropy gap” in generative models. At the core of this phenomenon lies a significant disparity between the probabilities of statistically frequent events and rare but meaningful ones. This “probability gap” is also reflected in Shannon's information entropy. Low entropy indicates a distribution concentrated around frequent patterns, resulting in high accuracy but little novelty (i.e., little information). High entropy broadens the space of possible outcomes, generating highly informative, “creative,” but often false or hallucinatory responses.

Some AI researchers have suggested that the available training data for AI models is already exhausted and propose continuing model development through the use of synthetic data generated by the models themselves. However, information entropy clearly indicates that such a data inbreeding approach would rapidly lead to information degradation and model collapse. This phenomenon is also known as model autophagy disorder (MAD) [11, 12].

We argue that a possible solution to many of these fundamental limitations lies in the creation of an adaptive open system based on mechanisms that take into account the physical nature of information and computational processes. This requires abandoning computations based on abstract symbols from an internal system alphabet and the statistical construction of complex separating hypersurfaces. Instead, it calls for feature generalization based on the formation and transformation of the system’s internal energy structure. Unlike traditional approaches, our concept regards information as a structure of subjectively measured parameters of external energy, which forms the system’s internal energy landscape — a model of the external world. Information controls the processes of self-organization and system evolution by minimizing its internal energy, forming a set of interconnected local energy attractors corresponding to individual classes or patterns. This approach is consistent with the principles of construction and functioning of both physical and biological systems, where external energy is transformed into internal energy during the stages of sensory perception and processing, thereby governing the dynamics and structure of the system.

Such a perspective makes it possible to bridge information theory, physics, and neuroscience, offering a unified framework for understanding complex systems. Overall, this research is intended to bridge the existing gap between theoretical physics, which views computation as a physical process involving thermodynamic costs, and computer science, which treats it as the manipulation of abstract symbols within a system's internal alphabet.

Energy Landscape of the Information System

The idea of converging the understanding of the general nature of energy and information can be realized by applying the concept of the energy landscape to both energetic (thermodynamic, dissipative, quantum, etc.) and informational, biological, economic, or social systems.

In physics, an energy landscape is a concept that describes how a system evolves in state space, striving toward a state of minimal energy. For example, in mechanics, an energy landscape can be a representation of a system's potential energy as a function of its spatial coordinates, while in thermodynamics, it reflects free energy as a function of system parameters (such as temperature, pressure, concentration, etc.). In such landscapes, the minima—either local or global—correspond to stable states of the system, while maxima and saddle points represent unstable states. The differences between global and local minima determine the complexity of the system's dynamics or evolution.

The concept of an energy landscape is widely used in chemistry and biophysics (e.g., in protein folding and catalysis), as well as in statistical and quantum mechanics. Undoubtedly, real physical systems possess a complex energy landscape that describes the system's energy across various states.

In the general case, the total energy of a system, U_{total}, includes both the local energies of i-th structural elements (units) of the system $U_{local,i}$, and the interaction energies between units i and j, denoted as $U_{bond,ij}$:

$$U_{total} = \sum_{i} U_{local,i} + \sum_{i<j} U_{bond,ij}. \qquad (1)$$

The energy landscape describes the dependence of the total energy U_{total} on the configuration of the system. The bonding energy U_{bond} characterizes the interactions between system elements and plays a key role in shaping the energy landscape. Examples of such interactions include: covalent or ionic bonds in chemical compounds, Van der Waals intermolecular forces, elastic interaction energy in solids (e.g., in crystals or molecular lattices), and the energy of magnetic or electric dipoles in systems with magnetic moments or polar molecules. These interactions are often nonlinear, and their energy depends on the distance between structural points.

Fig. 1 presents the general structure of the system's energy landscape, which is described by equation (1). In the general case, *a stable internal energy landscape El* of an open energy system is defined as the distribution of internal energy across structural elements (units) U_i and the connections between them W_{ij}, without considering the processes of irreversible heat dissipation ΔT, which are associated with energy fluctuations of the units around their stable state, and without accounting for external energy or energy fluxes between the system and the environment.

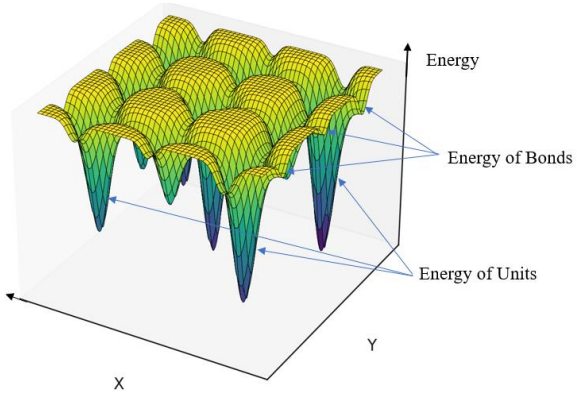

Fig. 1. General structure of the energy landscape of the system

$$El_{rest} = min(\sum_{i=1}^{N} U_i + \sum_{i=1}^{N}\sum_{j=1}^{N} W_{ij}) = Tr1\,. \qquad (2)$$

In this formulation, we refer to this energy landscape as *the landscape of the system at rest or the energy equilibrium landscape* El_{rest}, and the total energy of the system in this state as *the rest threshold* $Tr1$. In this state, the system reaches *maximum entropy* for a given macrostate, where no energy, mass, or substance flows are observed. For example, in thermodynamic equilibrium, all thermodynamic forces, such as temperature, pressure, or chemical potential gradients, are zero, and all internal processes that could occur within the system (e.g., chemical reactions, heat transfer) are complete.

Even if energy fluctuations at the microscopic level continue—such as changes in internal energy parameters (e.g., particle motion)—the system remains in an *equilibrium (rest) state*. This state defines the maximum number of possible microstates for a given macrostate that do not lead to its change. In other words, when the structure stabilizes and its internal energy is minimized, we describe the system as having reached its maximum entropy level for a given macrostate.

The rest or equilibrium state of a system is an idealized concept for a closed thermodynamic system. When such a system reaches a state of rest, it possesses maximum entropy. If we consider an open system that continuously receives external energy, allowing it to maintain a non-equilibrium state and create ordered structures, such a system—referred to as a dissipative system—experiences *a local decrease in its internal entropy due to structuring, leading to the formation of dissipative structures*. At the same time, the overall macrostate of the system may remain unchanged, which still results in an increase in total thermodynamic entropy, in accordance with the second law of thermodynamics. These processes form the foundation of Prigogine's theory of dissipative systems [13].

Examples of systems in a rest state, depending on the interpretation of energy, include: a gas in a closed, isolated vessel that has reached thermodynamic equilibrium (where temperature, pressure, and density are uniform throughout the volume); a molecule in vacuum; an isolated biological neuron in a resting state; an artificial neural network in a resting state (i.e., in the absence of input signals and changes in internal parameters).

Analogies (metaphors) can also be drawn for economic or social energy at rest, such as a company's financial reserves stored in a bank or a society's system of cultural and moral values.

However, it is important to note that this state indicates a fundamental limit—*the rest energy threshold of the system, which defines the minimum internal energy of the system and the maximum entropy* for a given structure and a specific set of macrostates. The energy landscape of a system at rest depends on its internal physical structure. Thus, this landscape forms the energetic foundation of the system, which determines its resilience to external energetic influences.

The minimum energy of the rest-state landscape is ensured by *minimizing the physical structure of the system and/or minimizing the internal (potential) energy of the units and their interconnections*. Structural minimization can be associated with its dynamic adaptation in response to changes in external energy influences or with evolutionary structural changes, as observed in bifurcation points of dissipative systems or in biological systems. However, if we speak about minimizing the internal energy of the units and their interconnections, we must recognize that *this energy itself has a structure*. Examples of describing such energy structures can be found in quantum mechanics and quantum chemistry, such as Schrödinger's equation, the Hartree-Fock method, the Lennard-Jones potential, and density functional theory (DFT), among others.

From quantum mechanics, the bond energy E_{bond} reduces the total energy of individual atoms E_{atoms} within the overall molecular energy structure $E_{molecule}$:

$$E_{bond} = \sum E_{atoms} - \sum E_{molecule}. \qquad (3)$$

This implies that part of the potential energy of the units is used to create interconnections between them. Thus, the energy landscape of the system at rest consists of "wells" of potential energy, represented by the units, and "channels" for potential energy transitions between wells, represented by interconnections (Fig. 1). The potential energy within these wells and channels has an internal structure, consisting of energy levels or sublevels. These potentials effectively determine the energy capacity of the system and indicate the "bottom level" of this capacity relative to some reference point (the "plateau" of the landscape). For example, in an atom, the potential energy of an electron in the nucleus's field defines the "bottom" of its energy state relative to the nucleus's energy. Metaphorically, we can visualize this level as the geodetic height of a point in an energy landscape above sea level.

Thus, the formation of stable structures leads to a decrease in the internal energy of the system in a resting state.

If we apply this description of the rest-state energy landscape to ANNs, we see that:

1. For most types of models, the rest-state energy landscape is static and, as it might seem, consists solely of weight coefficients, which we can interpret as the energies of interconnections between neurons (units). In these models, the energy is considered in only one direction for signal propagation. However, this is not the case, because in ANNs, the interconnections exist "physically," but they do not exist as distinct energy elements. If we look at the McCulloch-Pitts model, which forms the foundation for most neural models in ANNs:

$$Y = f\left(\sum_{i=1}^{n} w_i x_i + Q\right), \qquad (4)$$

where Y is the neuron's output (response), f is the activation function, w_i is the weight coefficient of the i-th neuron input, x_i is the input signal to the i-th input, Q is the activation threshold of the neuron. When, we see that the weight coefficients serve as modulators of the input signals (input energy) of the neuron, rather than independent energy elements of interconnections. However, if we consider that the input signals correspond to the values of bits (as in the classical model), then from an energy perspective, the sum of the weight coefficients could determine the energy structure and the possible depth of the potential "well" of the unit's energy landscape in the ANN and, consequently, the overall energy landscape of the system in its resting state (2). That is

$$U_{jrest} = \sum_{i=1}^{N} w_{ij}\,; \qquad (5)$$

$$El_{rest} = \sum_{j=1}^{N} U_{jrest} = \sum_{i=1}^{N}\sum_{j=1}^{N} W_{ij}, \qquad (6)$$

where U_{jrest} is the energy of the j-th unit (neuron) in the resting state.

Indeed, if the input $x_i = 0$, then this input energy does not leave the i-th sublevel of the unit's potential "well," with $w_{ij}x_i = 0$. However, if $x_i = 1$, the energy output from this sublevel equals w_{ij}, meaning that the sum of weight coefficients in this case determines the maximum "depth" of the potential "well."

In general, drawing an analogy with an energy system, such as an atom, we could compare the operation of a neuron to the ionization process of an atom, where the linear combination of inputs $\sum w_i x_i + Q$ models the release of an electron from the atom, i.e., the generation of free energy. Meanwhile, $w_i x_i$ models the transition of an electron to another energy sublevel, where w_i defines the energy of the sublevel in the resting state. However, this is merely a conditional analogy that, in this case, lacks a common energetic foundation.

Thus, the formal neuron model is an informational model based on the mathematical, or more precisely, statistical abstraction of information processes and, in its classical interpretation, is far from describing thermodynamic processes.

2. The values of individual weight coefficients in most ANN models could be interpreted as a conditional internal energy of the system, which lacks structure and is not connected to the system's input energy. Indeed, each neuron simply distributes (duplicates) its output energy evenly across all neurons in the next layer. This energy distribution can be compared to its dissipation. However, in reality, such "dissipation" in an ANN (if one considers its physical implementation) is associated with additional energy costs required to maintain the energy level of the neuron's response at the inputs of multiple neurons in the next layer.

The weight coefficients of inputs are associated with a different type of energy—pseudo-energy. However, this pseudo-energy does not govern the

distribution of external energy across the system's internal energy landscape, as it might seem. Instead, it determines the coordinated potential contributions of individual elements of the neurons' input energy, directed at overcoming their activation thresholds. This pseudo-energy represents the error energy, which is minimized during the iterative learning process.

Thus, the weight coefficient formally defines the "correct" contribution of input energy to overcoming the activation threshold, from the perspective of minimizing error energy. In other words, it determines the release of energy from the "potential well" to achieve the required response of both the neuron and the ANN as a whole.

3. The input energy of an ANN is normalized and constrained by a representation system, which can conditionally be compared to sensory perception that segments and transforms external energy into internal energy. Thus, both input and output energies are represented as another type of pseudo-energy, *expressed in the symbols of the system's internal alphabet*, for example, in binary or decimal numerals.

Such a representation of input energy, along with other specified constraints, leads to the formation of a deterministic set of possible *microstates of the system*. In ANNs, these microstates are determined by the sets of weight coefficients in the resting state and the distribution of input energy across the network, i.e., the responses of neurons during information processing.

The set of *macrostates* of an ANN is then defined by the responses of the output layer neurons and, from an energetic perspective, forms the system's free energy. In this sense, the network's output reflects a new stable state of the system (in which it has minimized its energy), conditioned by the interaction of input energy (the input signal vector) with the energy landscape.

One could say that an ANN in the inference state (the operation of a trained network) is in a state of *stable energetic nonequilibrium*. In this state, the system exits its resting state, but the structure of its internal energy landscape remains unchanged. The system maintains a stable internal (physical) structure and dynamics, as well as predefined macrostates.

This state is associated with the existence of a second fundamental threshold—the system's activation threshold *Tr*2. This threshold defines the system's internal metric, allowing for the determination of the depth of "wells" and "channels" in the system's energy landscape. The magnitude of a unit's local activation threshold determines the *depth of its potential energy "well," i.e., its energy capacity, and depends on the internal structure of its energy*.

For example, at the quantum level, an electron must overcome several energy sublevels to detach from an atom, meaning it must surpass the ionization barrier, which can be considered an analogue of the activation threshold. Different systems will have different activation thresholds: for biological neurons, it is the action potential; for an artificial neuron, it is the activation function. In dynamic systems, the activation threshold may be described by Lyapunov function or functional. Depending on the type of system, the complexity of interactions, and system properties, other methods may be used, such as Jacobian matrices, phase diagrams, energy barriers, feedback mechanisms, statistical methods, critical phenomena, and more. Thus, the type and structure of internal energy define the rules or conditions for the system's activation. In the general case:

$$E_{act(i)} = \left(U_{min(i)} + \Delta U_i\right) > Tr2_i; \quad (7)$$

$$\Delta U_i = F(X). \quad (8)$$

where $E_{act(i)}$ – activation energy of the *i*-th unit, $U_{min(i)}$ – potential resting energy of the *i*-th unit, ΔU_i – additional energy of the *i*-th unit, $Tr2_i$ – activation threshold of the *i*-th unit – the limit of its potential capacity, $F(X)$ – function of input energy distribution across units.

Let us imagine a surface with pits connected by hollows or channels, and possibly hills. If we pour water onto this surface, we will observe the process of energy distribution across the energy landscape of the system, filling the pits and channels. This representation vividly illustrates the process of energy distribution across the energy landscape of the system.

If there is an influx of energy into the system and no effective outflow, the energy will overflow the system's capacity, and the energy connections between the units may be destroyed, as their potential capacity is much smaller than that of the units themselves. This connection capacity determines the third fundamental threshold of structural stability or system resilience *Tr*3.

$$E_{str(ij)} = \left(W_{min(ij)} + \Delta U\right) \leq Tr3_{ij}, \quad (9)$$

where $E_{str(ij)}$ – energy of the stable connection between units *i* and *j*, $W_{min(ij)}$ – minimum potential energy of the connection between units *i* and *j*, $Tr3_{ij}$ – threshold of structural stability for this connection.

If the destruction of connections exceeds the structural stability threshold of the entire system, this may define the bifurcation point of the system, i.e., the point of its state change or further development, as well illustrated in Prigogine's theory of dissipative systems.

In an ANN, the activation threshold is represented by a rule or function that defines the artificial neuron's response. In most modern neuron models, this activation function (e.g., sigmoid or ReLU) is related to response normalization and does not reflect the level of threshold exceedance. However, it can serve as an analog of the neuron's potential energy level in the activated state.

The existence of two thresholds, *Tr*1 and *Tr*2, determines the stability of any system against fluctuations within the inter-threshold space. Exceeding the activation threshold by a certain magnitude results in the same stable system response, for example, at the system's output. To transition the system into an unstable state, external energy is required to alter the resting energy, the activation threshold, or to surpass the structural stability threshold.

Since the energy landscape consists of multiple components, an energy fluctuation that disrupts local elements of the *Tr*3 threshold may not lead to the complete collapse of the entire landscape. However, an increase in the number of such fluctuations can bring the system to a bifurcation point. This approach to analyzing state changes in the system's energy landscape offers a new perspective on system classification:

- *If the system can adapt to changing conditions by modifying its thresholds or internal energy parameters, it is classified as adaptive.*

- *If adaptation occurs through the restructuring of its resting energy landscape, which involves changing its internal structure, the system is evolutionary.*

- *If the system can create a new resting energy landscape by synthesizing its internal structure, it is classified as self-organizing.*

Thus, the activation threshold defines the maximum allowable fluctuation of internal energy for maintaining a resting (equilibrium) state. If this limit—determined by the structure of internal energy—is exceeded, the system transitions into a stable nonequilibrium state.

The bifurcation threshold depends on the system's structure, its resting and activation thresholds. If system response stability deteriorates with increasing fluctuations, this may indicate proximity to the bifurcation threshold.

The system may then transition either to adaptation through self-organization or to a state of degradation, meaning a nonequilibrium state with a disruption of its internal energetic and physical structure.

An important conclusion can be drawn: *the structure of energy serves as the foundation not only for shaping the system's internal energy landscape but also for its processes of adaptation, evolution, and self-organization.*

If we consider that system adaptation can occur not only through feedback but also through deep internal mechanisms of energy transformation, which involve changes in thresholds and the energy landscape, this could significantly expand system theory and cybernetics. Moreover, it could transform our understanding of learning in artificial neural networks, paving the way for a new perspective on their self-learning capabilities.

A look at the energy content of the Hopfield model and other energy-inspired ANN models

The first and most well-known ANN model that utilizes the concept of an energy landscape and an energy function is the Hopfield model [14, 15]. This model represents an attempt to bridge the gap between information theory and physics. Let us analyze how well-founded this convergence is. From the perspective of the energy states in the model, the most well-known physical interpretations of Hopfield's concept assert the following:

1. The energy landscape is interpreted as the set of all possible states of neurons.

2. The minimization (or rather, reduction) of energy during state changes, both in individual neurons and in the network as a whole, is illustrated by the Hopfield energy function.

3. A local attractor, which defines the "correct" state of a single neuron as an element of a global attractor, has a "basin of attraction". This basin is determined by the states of the neuron that converge toward a stable value, which corresponds to a stored pattern element. Failure of a neuron to reach the local attractor state is characterized as an undesirable phenomenon—a "trap" in a local energy minimum. This effect can arise due to interference between stored patterns.

4. Based on these assertions, it follows that there exists a set of global attractors with minimal energy, each of which corresponds to a stored pattern.

As a prototype for the Hopfield model, the Ising model is used to describe the properties of ferromagnetic materials [16]. The Ising model is a physical model, and its primary characteristics include energy, entropy, and temperature, which are used to describe the behavior of a system consisting of a large number of interacting spins in magnetic materials. In these materials, interactions between magnetic moments (spins) lead to complex and disordered states.

Spins represent the orientation of atomic magnetic moments and are associated with the energy of their interactions. The Ising model was developed to understand how local interactions between neighboring spins can lead to global ordering in systems, such as the phase transition of a ferromagnet. This ordering depends on temperature: at low temperatures, spins tend to align in the same direction (ordered state), and at high temperatures, spins become randomly oriented (disordered state). This model does not account for all forms of energy, nor does it describe energy dissipation or free energy formation, meaning it is not a complete thermodynamic model.

If we analyze the model in more detail, it studies the *collective behavior of spins S_i, which can take values* ± 1. The interaction between spins is described by a matrix of constant or random values J_{ij}. This J_{ij} can be interpreted as t*he conditional interaction energy, which can reflect spin correlation: a positive value corresponds to parallel spin orientation, or a negative value corresponds to spin misalignment (disorientation)*.

It is crucial to note that J_{ij} represents the interaction energy rather than the internal energy of a single unit (atom). However, in the classical Ising model, all J_{ij} values are set to +1, meaning it is axiomatically assumed that spin states tend to converge toward a single stable state with parallel alignment. The current states of spins depend on temperature, which influences the probability of spin reorientation according to the Boltzmann distribution: the lower the temperature, the higher the probability of parallel spin alignment. During the dynamic change of spin states, a complex energy landscape of the system is formed. The dynamics of the Ising model are described through the energy functional, the Ising Hamiltonian H:

$$H = -\frac{1}{2}\sum_{i \neq j} J_{ij} S_i S_j - \sum_i h_i S_i, \qquad (10)$$

where h_i is the external field.

As the system evolves, its energy decreases until it reaches a minimum, characterizing a stable state with parallel spin orientations. In the absence of an external field ($h_i = 0$), this state is fully defined by the matrix of spin interaction weights. Thus, i*n the stable equilibrium state of the system, the energy landscape, according to equation* (10), *is represented by the weight matrix with values of* +1, reflecting the state of the system with minimum energy and maximum thermodynamic entropy.

The Ising Hamiltonian is a discrete analog of the Lyapunov function, which is used to describe continuous stable states of dynamic systems [17]. In the Hopfield model, the Ising model's concept is adapted to solve the

problem of building associative memory. The energy function of the Hopfield model, which decreases with each step of neuron state updates, is described by the following functional:

$$E = -\frac{1}{2}\sum_{i\neq j} w_{ij}s_i s_j - \sum_i \theta_i s_i, \qquad (11)$$

where s_i are the neuron states (+1 or -1), w_{ij} are the connection weights between neurons, and θ_i are the threshold values.

One of the key differences between these models is that in the classical Ising model, the changes in spin states occur randomly using the Metropolis algorithm (a Monte Carlo method) [18], which models the relationship between the external and internal energy of the system. In contrast, in the Hopfield model, changes in neuron states occur deterministically based on comparing input signals with the signs of mutual correlations, represented by the sum of connection weights. The changes in neuron states aim to restore a specific state (pattern) due to the recurrent architecture of the model. That is, the architecture of recurrent connections in the Hopfield network acts as a sort of "broom" to maintain its dynamics.

In the Ising model, the threshold for the spin "flip" depends on the temperature T and the change in energy ΔE. This threshold is defined by the probability P with which the system "decides" whether to adopt a new state that might have a higher energy:

$$P(\Delta E) \propto e^{-\Delta E/k_B T}, \qquad (12)$$

where k_B is the Boltzmann constant.

The probability P, according to the Boltzmann distribution, serves as an analog to the energy threshold that determines the change in the spin state. *Therefore, the relationship between the energy threshold and thermodynamic parameters defines the physical basis of the Ising model.* We can see that in the Ising model, this threshold is dynamic, while in the classical Hopfield model, the threshold θ_i for changing the state of the neuron is static and equal to 0. *This threshold, along with other parameters of the Hopfield model, is not related to thermodynamics, which defines the Hopfield model as informational.*

Let us pay attention to the fact that the Hopfield model uses the concept of the system's "energy landscape" as the set of the neurons' states, but does not provide its definition for the equilibrium resting state. However, if this is an energetic system, then there must exist a landscape in the resting state.

It is easy to see that, according to the Hopfield energy function, the energy landscape of the system in the resting state is formed by the sum of the weight coefficients at the neuron inputs. In other words, the weight coefficients acquire the meaning of *the internal energy of the network in the resting or equilibrium state*. Indeed, according to the Hebbian learning rule, these weight coefficients are formed once during the learning process and are not dynamically updated afterward:

$$w_{ij} = \frac{1}{N}\sum_{k=1}^{p} x_i^k x_j^k, \qquad (13)$$

where x_i^k is the response value of the i-th neuron in the k-th sample, N is the number of neurons in the network, and p is the number of patterns to be memorized. Thus, in the Hopfield model, the weights w_{ij} are formed based on the correlation of patterns that the network must memorize, rather than on the basis of fundamental energy interactions.

Then, according to formula (11), the minimum energy of the network in the equilibrium state E_{min}, i.e., theoretically, if all the states of the neurons (their responses) were perfectly correlated with each other based on their weight coefficients, will be:

$$E_{min} = -\frac{1}{2}\sum_{i\neq j} w_{ij}, \qquad (14)$$

Thus, the network has *one global minimum energy in the resting state*, which is a theoretical limit, and practically, in a working network, this is not achievable due to the various types of correlations for different signals and the peculiarities of the weight coefficient formation algorithm. One can conclude that, indeed, for each memorized pattern, there will be its own minimum value of the energy function, which, in practice, will depend on the degree of distortion of the memorized patterns. This can be interpreted as *the existence of global energy minima for each memorized pattern*.

This is explained by the fact that the signs of the sums of the weight coefficients of a neuron indicate the "correct" or "desired" correlation of the input signals with the memorized patterns—either positive or negative. The sum of the weighted input signals of a neuron can be interpreted as a "*correlational vote*," as a result of which a decision is made to change or maintain the current state of the neuron. The result of this "voting," reflecting the maximum possible correlation of signals, will depend not only on the values of the weight coefficients but also on the current states of the neurons.

If complete correlation according to the weight coefficients is not achieved, then the system will not reach the desired global minimum corresponding to the memorized pattern, which will be reflected in the energy formula as an undesirable phenomenon of a local energy minimum. Therefore, in the Hopfield model, *the energy function can increase at the level of individual neurons if local correlations are disrupted*. However, the global dynamics of the network, taking into account possible constraints, will aim to reduce the overall energy function. This emphasizes that the Hopfield model works with a global energy function, and local deviations in individual elements can be compensated by the overall behavior of the system.

Thus, the following conclusions can be drawn:

1. The Hopfield model, unlike the Ising model, is informational, meaning it is abstracted from any thermodynamic or other energetic interpretation. Therefore, when considering it, it would be more appropriate to use the term "conditional" or "pseudo" energy. However, the use of concepts such as "energy landscape" and "energy function" still represents powerful conceptual steps. *The minimum of the Hopfield energy function demonstrates the attainment of a stable state and a reduction in the system's informational entropy*.

2. The classical Hopfield model has a spatially linear energy landscape of the system in its resting state,

represented by the sum of the weight coefficients of the neurons. This is explained by the physical structure of the network. In the Ising model, this landscape reflects the lattice structure of atoms. When considering the Hopfield model, a dilemma arises regarding the interpretation of the energy of the weights: they could be attributed to the elements of the internal energy of neurons (Equation (2)) or to the energy of neural interactions (Equation (6)), since they are a function of the reference values of their internal states according to Equation (13). However, due to the structure of the McCulloch-Pitts model (Equations (4), (5), (6)), they should still be attributed to the elements of the internal energy of the units (neurons). In this case, the coefficient ½ in the Hopfield energy formula (Equation (11)) is redundant, unlike in the Ising Hamiltonian (Equation (10)), where the weight coefficients indeed represent the interaction energy of atoms.

3. The Hopfield energy function is separate from the process of governing the network's dynamics and serves only as an illustration of the process. Therefore, the metaphor of "attractors," whether informational or energetic, is incorrect, as there is no energy or energetic state in the model that "attracts" other energetic or informational states. That is, the figurative representation of the network's operation as a ball moving along the surface of an energy landscape and falling into a "well" with minimal energy does not correspond to real processes.

This assertion is justified by the fact that the metaphor of an "element of a global attractor" could only apply to the sum of the weight coefficients of each neuron as the only explicit, rather than virtual, constant value. However, the problem is that this sum serves the same role for all "global attractors," and all correlations between "elements of the global attractor" are interconnected by the matrix of these weight coefficients. The decision to change the state of a neuron depends on the values of incoming signals, meaning the "attractor" must be dynamic and change depending on the input signals. It should be noted that in dynamical systems theory and nonlinear dynamics, there are concepts that describe such phenomena, but for this, a system must have: multistable states and a switching mechanism between them; or attractor metamorphosis with gradual or sudden changes in the shape or type of the attractor; or chaotic evolution of the attractor, among others, which is possible in complex systems with changing parameters or interactions. Therefore, it is more appropriate to consider not the metaphor of an "attractor" but the true correlational nature of these weight coefficients and their sum. This process can be called *correlational voting*, where the majority of weighted input signals "vote" for a positive or negative correlation of the output signal, i.e., for making a decision to transition to state +1 or -1.

4. If the energy function of the Hopfield model (11) indicates a decrease in conditional energy associated with a reduction in informational entropy, then the analogous energy function of the Ising model (10) also points to this. Thus, it can be concluded that *a physical spin-glass system produces not only thermodynamic but also informational entropy*. This is consistent with the theory of quantum informational entropy and Prigogine's theory of dissipative systems.

Overall, the functioning of the Hopfield model could be described more simply and accurately without using energetic terminology: the final stable state of the system depends on the initial state (input signal vector) and the correlation matrix defined by the weight coefficients of correlation. The process of state changes in the system is iterative and determined by recurrent connections. The final state of the system is characterized by achieving the maximum possible (stable) correlation of neuron responses (their states) in accordance with their weight coefficients. The energy function (Lyapunov function) in this interpretation takes on a demonstrative meaning and reflects the system's tendency toward a stable state or a reduction in internal informational entropy.

However, if we decide to consider the operation of this network purely from an energetic perspective, let us note that the internal energy of the network, as we have established, corresponds to the energy of neurons (units) according to formulas (5, 6), whereas the energy of neuron interconnections remains undefined. Suppose the system "does not know" what a bit of information is and instead treats it as a value of potential energy that needs to be "disposed of" in some way: stored, distributed, dissipated, or converted into output energy. Then, the dynamics of neuron state changes are determined solely by elements of external energy entering the system. In our case, this energy corresponds to the symbols x_i of the bipolar alphabet (+1 – maximum energy, or -1 – minimum energy). Indeed, if no signals are supplied to the network's input, then the output of each neuron will either be zero or contain random energy values.

Thus, the values of the internal energy of the network, expressed in the form of the weight coefficient matrix w_{ij}, define the threshold values of the system's resting state. The dynamics of state changes in the system are determined by the excess of external energy over these resting threshold values within the threshold activation value θ_i. This process should be viewed not as the reduction of the system's internal energy to a minimal (stable) level but rather as its redistribution among local points of the energy landscape (units).

An important feature of the model is that the energy landscape, determined by the weight coefficient matrix, is formed based on samples of external energy that need to be memorized, rather than in the form of conditional error energy, as in a multilayer perceptron, for example. Thus, the energy landscape of the Hopfield model could be considered as an *internal model of the external world perceived by the system*. Thus, the process of iterative redistribution of external energy across the local levels of the energy landscape is represented in the form of a Lyapunov function, expressed through the Hopfield energy function (11). The result of this redistribution indicates the level of external energy exceeding the internal resting-state energy at the local extrema of the landscape. After comparing this excess with the threshold value θ_i of the functional, it is reflected in the system's output signals. Therefore, θ_i serves as the decision threshold for redistributing external energy among the network elements.

According to Hopfield's formula (11), with each step of energy redistribution, the total conditional energy

E of the network decreases, which, in accordance with the laws of thermodynamics, reflects an increase in the system's thermodynamic entropy. That is, external energy diminishes as it is "dissipated" among neurons and fills the "wells" of potential energy in the units (neurons) until a state is reached where further changes become impossible, and the system stabilizes.

Thus, we could also conclude that the Hopfield energy formula represents the dynamics of entropy change, inherently linking informational and thermodynamic entropy. However, a drawback of this description is that neurons must "dissipate" the accumulated potential energy in their "wells" before the next state update cycle. As a result, to maintain dynamic operation, the system will continuously require additional external energy.

One of the conceptual problems of both the classical Hopfield model and this alternative energetic description of its operation is the absence of a self-halting mechanism. The system itself will continue running indefinitely without state changes (in an update loop) until an external stop signal is received.

Thus, the Hopfield network, as both an energetic and informational system, can be classified as a system with *a deterministic static structure, a deterministic static generalized energy landscape in the resting state (formed instantly rather than iteratively), and a deterministic dynamic energy landscape during the process of minimizing informational entropy*.

Other well-known energy-based neural network models, such as Restricted Boltzmann Machines (RBM), deep energy-based models, diffusion models, quantum neural networks, Energy-Based Transformers, or Latent Variable Energy-Based Models (LV-EBMs) of LeCun and his followers [19–21], are even further from describing physical processes than the Hopfield model. All these models rely on the statistical paradigm of learning. However, once we speak of the statistical nature of learning associated with some artificially introduced (pseudo) energy, this implies that for analyzing the statistics influencing the global behavior (reaction) of the system, we need an algorithm external to the model itself—for example, the back-propagation algorithm, which serves as a kind of analogue of Maxwell's demon in thermodynamics. Such an approach is not physically/biologically justified, and one may conclude that all modern energy-based neural network models are not genuinely energy-based from a physical/biological perspective, since they are not grounded in the consideration of the system's internal energy and its dynamics (minimization) during the transition to equilibrium, but instead in the optimization of external functions employing pseudo-energy (loss functions).

The same problem arises in Friston's Predictive Coding Theory [22, 23], which employs the principle of free-energy minimization, but this "free energy" is in fact an abstract quantity related to a probabilistic (Bayesian) distribution, rather than to physical energy.

Energetic interpretation of the formal neuron model

The McCulloch-Pitts (MCP) neuron model is a simplified mathematical representation of a biological neuron, which, nevertheless, has become widely used and has formed the basis for the construction of modern ANNs. This informational model is implemented algorithmically; however, modeling it as an energy system will allow us to understand the limitations of this model compared to a real physical system, which a biological neuron undoubtedly is.

Let us consider the MCP model as a single-layer perceptron in its classical version with binary signals {0, 1}, weight coefficients in the range {−1, +1}, and a Heaviside activation function. The energy model of the MCP can be examined based on the construction of its energy landscape. We have already mentioned that the energy landscape of an MCP neuron in its resting state is represented by a set of weight coefficients that determine the energy contribution of each input to overcoming the "potential energy well," the depth of which is limited above by the activation threshold *Q* (4-8).

Additionally, we have noted that the energy of the MCP model represents only the energy of a single unit in the energy landscape of a more complex system (e.g., an ANN). This unit (neuron) does not form energy interactions with other units, unlike real physical systems (1).

Fig. 2 shows the general structure of the MCP energy model, where $X = (x_1, x_2, \ldots, x_n)$ is the set of input signals representing the system's input energy; Y is the response of the MCP model, representing the system's output energy; $W = (w_1, w_2, \ldots, w_k)$ are the weight coefficients representing the energy contributions of the inputs to overcoming the activation threshold; Σ is the summator of the weighted inputs, representing the current depth of the system's (unit's) "potential well"; and f is the Heaviside activation function, representing the upper threshold of the "potential well."

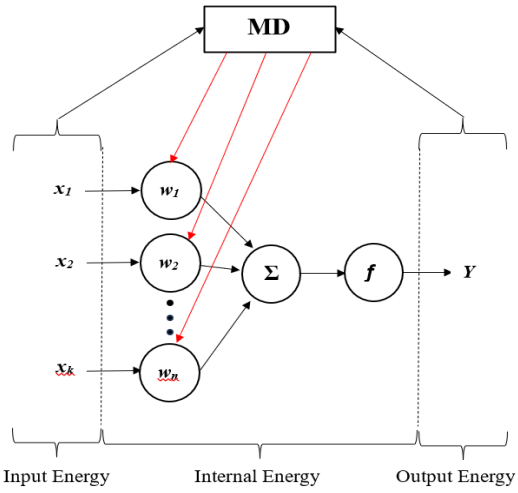

Fig. 2. General structure of the energy model of the MCP

The sum of the weighted inputs represents the internal energy of the system. This energy is divided into excitatory (activation energy), where the weight coefficients have a positive sign, and inhibitory (suppressing) energy, where the weight coefficients have a negative sign.

Thus, if the "positive" total input energy exceeds the "negative" or inhibitory energy, the activation threshold $Q = 0$ is surpassed, and the system, using the activation function, decides to generate an output signal $Y = 1$, regardless of the activation potential V:

$$V = \sum_i w_i^+ x_i + \sum_i w_i^- x_i = V^+ + V^-, \qquad (15)$$

where w_i^+ are the positive weight coefficients that determine the positive activation potential V^+, w_i^- are the negative weight coefficients that determine the inhibitory potential V^-.

Thus, the inhibitory potential V^-_{rest} determines the "depth" of the potential well and the minimum energy of the unit in the resting state V_{rest}:

$$V_{rest} = V^-_{rest} = \sum_i w_i^- . \qquad (16)$$

This "depth" of the potential well is represented in the energy landscape of the unit model (formal neuron) in the resting state (Fig. 3).

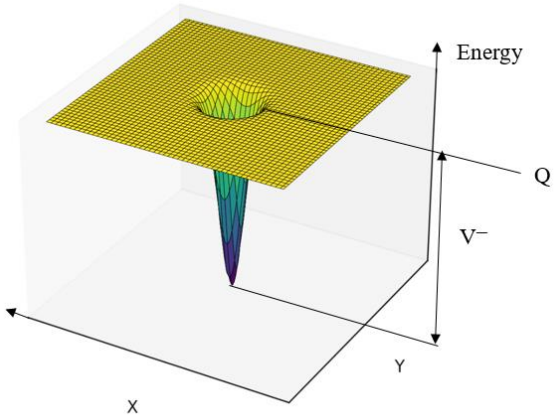

Fig. 3. Energy landscape of the MCP model in the resting state

The process of normalizing the output energy of the system (representing it as a single-bit value based on Landauer's principle) can be associated both with the "dissipation" of energy exceeding the normalized value of "1" and with the use of additional energy in cases where $V < 1$.

If $|V^-| > |V^+|$, then all the energy V^+ will be dissipated with energy release in accordance with Landauer's principle [24].

The model in Fig. 2 also presents "Maxwell's Demon" (MD) [25]. This is an abstract thermodynamic concept proposed by the Scottish physicist James Clerk Maxwell, which is widely used in information physics and quantum computing. In this case, we consider it as an analogy to an external energy system relative to the given system, which implements the learning algorithm of the neuron. As an energy system, MD regulates the flow of input energy for its transformation into the internal energy of the system by purposefully adjusting the weight coefficients.

Practically, MD minimizes the error energy at the system's output. This error energy represents "pseudo-energy" relative to the system's internal and external energy.

To consider the MCP model as a thermodynamic system, we must also define the system's microstates and macrostates. We define the system's microstates as the set of weighted inputs, which correspond to specific system outputs—macrostates. In this model, there are only two macrostates: $Y = \{0,1\}$, where only the macrostate $Y = 1$ corresponds to neuron activation.

The probability of neuron activation P_{act} can be considered as a conditional probability:

$$P_{act} = \sum_{V^-} P(V^-) P(V^+ \geq (Q + V^-) | V^-), \qquad (17)$$

where $P(V^-)$ is the probability that the inhibitory inputs sum to V^-; $P(V^+ \geq (Q + V^-)|V^-)$ is the conditional probability that the excitatory inputs sum to a value exceeding the effective threshold $(Q + V^-)$.

This reflects the influence of inhibitory contributions: the "deeper" (i.e., the larger in magnitude V^- is), the higher the effective threshold for excitatory inputs, and thus, the lower the probability that the condition $V^+ \geq (Q + V^-)$ will be satisfied. If we define the number of microstates leading to activation as Ω_{act}, then the thermodynamic entropy S_{therm} is given by:

$$S_{therm} = k_B \ln \Omega_{\text{act}}. \qquad (18)$$

The number of microstates Ω_{act} depends on how many combinations of weighted excitatory and inhibitory inputs satisfy the condition $V^+ \geq (Q + V^-)$. That is a "deeper well" (stronger inhibition) reduces Ω_{act} because more excitation is required to overcome the threshold, leading to a decrease in thermodynamic entropy for this activation macrostate. Similar reasoning applies to the second macrostate $Y = 0$, where the number of microstates that do not lead to activation, $\Omega_{\text{non-act}}$, is determined by the condition $V^+ < (Q + V^-)$. Assuming an equiprobable distribution of all 2^N microstates, where N is the number of inputs, the probability of activation can be expressed as:

$$P_{act} = \Omega_{act} / 2^N, \qquad (19)$$

then

$$\Omega_{act} = 2^N P_{act}. \qquad (20)$$

In this case, the thermodynamic entropy for the activation macrostate S_{act} is given by:

$$S_{act} = k_B \ln(2^N P_{act}) = k_B (N \ln 2 + \ln P_{act}). \qquad (21)$$

Thus, as P_{act} increases, the thermodynamic entropy also increases for a fixed value of N and an equiprobable distribution of microstates.

Similar reasoning applies to the determination of the thermodynamic entropy for the second macrostate $S_{non\text{-}act}$. However, during the learning process of a neuron, we restrict the number of permissible input combinations, thereby making them non-uniformly distributed for a given (defined) macrostate. That is, for example, to activate a neuron, we apply the conditional probability given by equation (17). In this case, we use only a significant subset T from the total set of 2^N combinations:

$$|T| \ll 2^N. \qquad (22)$$

If we assume that the distribution of combinations within T is uniform, then:

$$P_{act} = \frac{\Omega_{act}}{|T|}, \qquad (23)$$

which means that maintaining the same probability P_{act} under conditions of a non-uniform distribution of input combinations – leading to the selection of a significant subset T – requires a smaller number of effectively active input combinations and, consequently, a smaller number of possible microstates Ω_{act}. The same reasoning applies

to the subset of combinations *T'* that do not lead to neuron activation.

Thus, we observe that during the learning process of a neuron, its thermodynamic entropy as an energy system decreases for the selected significant subsets of input combinations T and T'. In contrast, for the overall equiprobable distribution of input combinations, the thermodynamic entropy for both macrostates increases as the number of microstates grows to 2^N.

Let us consider the informational (Shannon) entropy *H* of the MCP model:

$$H = -[P_{act} \log_2 P_{act} + (1 - P_{act}) \log_2 (1 - P_{act})] . \quad (24)$$

We see that *H* reaches its maximum at $P_{act} = 0.5$, i.e., when the distribution of weighted excitatory and inhibitory inputs is equiprobable. This also corresponds to the *maximum thermodynamic entropy*.

When the conditional probability P_{act} *(equation 17) approaches 0 or 1, we observe a decrease in informational entropy, which also corresponds to a local decrease in thermodynamic entropy for the corresponding macrostate*.

Thus, if we define informational entropy through the probability distribution over microstates, it turns out that it precisely corresponds to thermodynamic entropy (considering the constant k_B). Moreover, the learning process, which leads to the selection of a significant subset of input combinations, modifies the distribution of microstates, thereby causing a local reduction in thermodynamic entropy.

Now, let us consider the output energy of the system as the Gibbs free energy *G* [26]:

$$G = H - TS , \quad (25)$$

where *H* is the enthalpy of the system (heat content, energy including heat), *T* is the absolute temperature, and *S* is the thermodynamic entropy of the system.

Gibbs free energy is a thermodynamic state function that characterizes the system's ability to perform useful work under isothermal and isobaric conditions (i.e., at constant temperature and pressure). This term is widely used in chemical thermodynamics to analyze the energetic aspects of chemical reactions and phase transitions. However, its interpretation as the output energy of a system is not commonly used. Generally, it indicates how much energy can be converted into useful work—mechanical, electrical, or other forms.

Importantly, Gibbs free energy can be linked to the internal structures of a system (e.g., crystal lattice, molecular organization). In this case, the value of *G* reflects how the system's structure influences its ability to perform work. Nevertheless, in machine learning and Bayesian statistics, the method of variational free energy is applied, utilizing thermodynamic principles for optimization [27]. Here, a functional analogous to Gibbs energy is introduced to minimize the difference between the model and the data.

However, if we consider that the free energy at the output of an information system "performs work" by creating an internal model of the external world for another system that perceives it, then the analogy between information generated at the system's output and Gibbs free energy takes on a completely different meaning and becomes well-justified. Furthermore, information enables systems to make decisions, predict, or change their state. For example, in control systems or robotics, the output of an ANN is used to control real physical objects, which serves as an analog to physical work. Then,

$$G = E_{str} = E_{total} - E_{unstr}, \quad (26)$$

where E_{str} is the structured part of the energy at the system's output. This portion of energy contains measurable parameters that reflect the order associated with the internal structure of the system. It is related to the reduction of informational entropy due to data structuring (e.g., classification, feature extraction, prediction, etc.); E_{total} is the total energy released by the system (enthalpy); E_{unstr} is the unstructured part of the energy. This portion of energy remains in a form that does not carry a stable structure in the system's output energy, which could otherwise represent information about the system's internal structure. This energy is analogous to thermal losses or noise, which increase thermodynamic and informational entropy.

We observe that in the MCP model, only for the macrostate "1" does the formation (release) of free energy occur, whereas for the macrostate "0," which signifies the absence of a neuronal response, all energy represents thermodynamic losses (*TS*) in accordance with Landauer's principle. If we disregard the energy costs associated with normalizing the unitary output of the neuron—additional energy for $0 \le V < 1$, or additional heat dissipation for $V > 1$—then it can be assumed that the inhibitory potential V^- makes a significant contribution to the formation of $TS = E_{unstr}$

$$TS = \Omega V^- E, \quad (27)$$

where *E* is Landauer's energy (7), defining the thermal energy released during the erasure of one bit of information; Ω is the total number of microstates; V^- is the inhibitory potential, determining the number of bits erased per microstate. Then, the total unstructured portion of the emitted energy, as the "chaotic" part of enthalpy, is equal to the total thermodynamic entropy of the MCP neuron:

$$S = \Omega V^- k_B \ln 2 = 2^N V^- k_B \ln 2 . \quad (28)$$

We observe that the total thermodynamic entropy of the MCP model increases with the number of microstates and the "depth" of the neuron's potential well (inhibitory potential). Consequently, as the total number of microstates in the system increases, the system's enthalpy grows, and its unstructured component E_{unstr} also increases, while the structured part of the output energy E_{str} reaches its maximum value for a limited number of microstates after learning. This corresponds to the minimum thermodynamic and informational entropy for the respective macrostates and remains unchanged thereafter. This conclusion aligns with the findings of Prigogine's theory of dissipative systems, which states that a system, when far from equilibrium, can "learn," meaning it transitions into a state that acts as an attractor in its dynamics. This implies that out of a vast number of possible microstates, the system "selects" those that optimally maintain order (low internal entropy) through energy dissipation. Thus, although the total entropy of the environment may continue to increase, the internal structure of the system stabilizes.

Let's clarify that in classical equilibrium thermodynamics, Gibbs free energy is minimized. However, here we are dealing with an open system which, considering "Maxwell's Demon," can only conditionally be regarded as a self-organizing system. In this case, the concept of the "maximum" of the structured energy E_{str} should be understood in terms of the system's optimal functional state rather than as a classical equilibrium state. This does not contradict the classical interpretation of Gibbs free energy, since we observe that as the system's total enthalpy increases, its total thermodynamic entropy E_{unstr} also increases, whereas the free energy E_{str} —or more precisely, its fraction as a component of enthalpy— decreases, along with its local thermodynamic entropy.

Overall, these conclusions are consistent with *the system's dynamic evolution equation*, which is analogous to the Langevin equation in statistical mechanics, describing the motion of a particle in a potential field with thermal noise

$$\frac{dx}{dt} = -\frac{\partial V(x)}{\partial x} + \eta(t), \qquad (29)$$

where x represents the internal energy of the system, $V(x)$ is the effective energy potential landscape, and $\eta(t)$ is a stochastic noise term modeling thermal fluctuations associated with thermodynamic losses [28].

In this equation, the function $V(x)$ serves as an effective energy potential, determining how the system tends to minimize its internal energy. Thus, $-\partial V(x)/\partial(x)$ defines the natural tendency of the system to reduce its energy, i.e., to transition to more stable states.

If a dissipation coefficient γ is added to the left-hand side of the equation, we obtain an equation for a dissipative system, which resembles the relaxation equation for energy in open systems.

General Conclusions:

1. In the MCP model, as well as in ANN models, information is represented as abstract elements of the system's internal alphabet, such as bits. These alphabet elements lack an internal structure that could be associated with energy structures, as seen in quantum physics or chemistry. However, informational bits are linked to energy according to Landauer's principle, allowing for an energetic model to be considered for both neurons and ANNs as a whole. In this case, the internal energy of the MCP model, as a unit of the energy system, can only be analyzed based on external "pseudo-energy," whose parameters (weight coefficients) are determined by "Maxwell's Demon" (an external training algorithm) through statistical observations. The statistical parameters of external "pseudo-energy" for determining the internal energy of a unit (MCP model) can be regarded only as "surface" properties of the system's energy, which do not fully reflect the structure of the unit's internal energy and cannot be used to form the energy of unit interactions, as occurs in physical systems.

2. During the training process of the MCP model, a stable, structured component of the energy emitted (generated) by the system is formed, which can be considered as Gibbs free energy. This structured energy exhibits a minimal thermodynamic and informational entropy for a limited number of effective microstates corresponding to certain (predefined during training) macrostates. The decrease in both entropies follows the same dynamics, reflecting the deep energetic nature of their interrelation. During training, the MCP model—more precisely, "Maxwell's Demon" as part of it—minimizes the external "pseudo-energy" in the form of the model's response error. This minimization is manifested in changes to the statistical parameters of the unit's internal energy, represented as weight coefficients. This leads to a reduction in the number of effective microstates and, ultimately, to the minimization of the system's internal energy.

3. Based on the presented interpretation of free energy, it can be stated that the parameters of structured energy at the system's output (for the MCP model, this could be a sequence of binary signals) reflect the internal energy of the system and, consequently, its internal structure, which can be interpreted as an *internal model of world perception*.

Energy interpretations of multilayer ANN models

To understand the energetic nature of information, let us consider multilayer ANNs, which can be regarded as ideal models not only for studying the principles of information processing but also for examining their relationship with the energy of physical systems. To this end, we will use the concept of an energy landscape and the energy model of the MCP previously discussed.

Fig. 4 presents the energy landscape model of a multilayer perceptron (MLP).

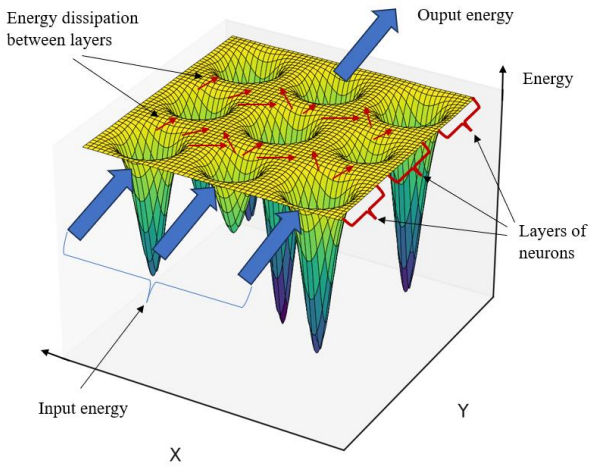

Fig. 4. The energy landscape model of MLP

This energy landscape consists of potential "wells" of units (MCP neurons) of varying depths in the resting state, forming the hidden layers. The model illustrates the general concept of energy convergence, which reflects the process of information convergence in an MLP when solving a classification task.

A key feature of this landscape is the absence of energy connections – channels for energy distribution between units. The output energy of units simply "dissipates" among the inputs of the next layer's units.

This results from the fact that the input weight coefficients do not generate connection energy but merely determine the depth of a unit's potential "well."

The process of output energy "dissipation" is represented by the fully connected architecture of MLP layers. In a real energy system, such "dissipation" would require amplification of the information signal at each neuron input. The lack of connection energy between structural elements in this energy model is a limitation compared to real energy systems (Fig. 1).

The coordinated formation of unit (neuron) energy in this ANN model (i.e., the formation of the system's energy landscape in its resting state) is also linked to the process of minimizing "pseudo-energy," governed by the system's "Maxwell's Demon" (e.g., the backpropagation algorithm). This process of coordinated unit energy formation, aimed at achieving the required set of macrostates, represents a trade-off between the generality of the extracted classification features and the accuracy of the classification task for a given set of classes. In many practical tasks, achieving this trade-off during training does not allow for a perfect solution. Based on the research results presented in Section 4, it can be concluded that the informational and thermodynamic entropies of a system whose energy landscape consists of multiple separate energy units (MCP models) will exhibit the same properties as an individual unit.

Evidence for these conclusions comes from studies by Ravid Shwartz-Ziv and Naftali Tishby (2017) et al. [29, 30], in which the authors proposed the Information Bottleneck (IB) theory. This theory examines deep neural networks, such as MLPs, and demonstrates that during training, there is a tendency for information entropy to decrease in the responses of hidden-layer neurons. Moreover, the reduction in information entropy is accompanied by a decrease in the variability of neuron responses in hidden layers, indicating a reduction in the number of microstates for the considered responses (macrostates) of the system. This also suggests a local decrease in thermodynamic entropy for these macrostates. For example, in the work of Shwartz-Ziv and Tishby (2017) [29], it is shown that in classification tasks on datasets such as MNIST, the information entropy of activation decreases by a factor of 2–4 from the input layer to the output layer, corresponding to a reduction in activation variance from 50% to 90% between the initial and deeper layers.

In IB theory, the authors introduce the concept of mutual information: $I(X; T)$ – the information that layer T contains about the input data X; $I(Y; T)$ – the information that layer T contains about the target data Y. Mutual information $I(X; Y)$ determines how much knowing one variable reduces the uncertainty of the other. In general, mutual information equals the sum of the entropies of the individual variables minus their joint entropy

$$I(X;Y) = H(X) - H(X|Y); \quad (30)$$

$$I(X;\mathrm{Y}) = H(Y) - H(Y|X); \quad (31)$$

$$I(X;Y) = I(Y;X); \quad (32)$$

$$I(X;Y) = H(X) + H(\mathrm{Y}) - H(X;Y). \quad (33)$$

If X and Y are independent, knowing Y does not reduce the uncertainty of X, then:

$$H(X|Y) = H(X) \; and \; I(X;Y) = 0. \quad (34)$$

If X and Y are fully dependent, knowing one variable completely determines the other, and $I(X; Y)$ reaches its maximum value, equal to the entropy of the variable with the smaller entropy.

These studies show that $I(X; T)$ decreases significantly toward the final layers, indicating a substantial reduction in the uncertainty of the input data. Moreover, the information entropy of activation $H(T)$ in each layer decreases during training. For example, in a 10-layer network, $H(T)$ in the last layer decreases by approximately 50-70% compared to the first layer. However, as expected, information entropy remains at the system's output, confirming the existence of a trade-off between the generality of extracted classification features—determined by the system's generalized energy landscape—and the stability (accuracy) of the system's responses to specific input information.

It has also been shown that, at the beginning of training, mutual information between a layer's representation and the input $I(T; X)$ in the initial layers can reach about 10–15 bits, meaning that a large amount of raw input information is preserved at these layers. As information propagates through deeper layers, $I(T; X)$ begins to decrease. For example, in the second hidden layer, $I(T; X)$ may drop to 8–10 bits. The most significant information compression occurs in the last hidden layers, where $I(T; X)$ can be reduced to 2–3 bits, while $I(T; Y)$ remains close to the entropy of the labels. This indicates that all relevant informational features necessary for decision-making are retained. These empirical results fully confirm our theoretical conclusions about the reduction of information entropy and the local decrease in the structured part of thermodynamic entropy at the output of the considered system (the energy model of an artificial neural network) for a specific set of its macrostates. Thus, since the considered system lacks energy connections between units, its internal energy is represented exclusively by the energy of the units themselves. Each unit perceives the energy of a neighboring unit as external energy, converting it into its internal energy using weight coefficients, which act as specific detectors of input energy. Consequently, the activation function of a neuron can be interpreted as a function that transforms internal energy into the output energy of the unit.

The structured external energy of the entire system, which represents the information at the output of the ANN, is formed by the activation functions of the neurons in the outer (final) layer. This system implements a process of parallel energy (information) perception and its convergence (compression) during training.

Let us call such an energy model of an ANN a first-type model. This type also includes convolutional neural networks (CNNs), whose main distinction lies in the segmentation of the input signal vector (input energy) and the local processing of individual segments.

The first-type ANN model can be classified as an energy system with: *a deterministic static structure, a deterministic static generalized energy landscape in the resting state, iteratively formed by the system's "Maxwell's Demon," a deterministic energy dynamic of units on the energy landscape, occurring during the minimization of informational entropy and the local minimization of thermodynamic entropy*.

We can also conclude that the generalized energy landscape of the system in the resting state, represented solely by the energy of units, along with the deterministic static structure of the system, imposes a conceptual limitation on its accuracy as the specified number of macro-states of the system increases.

In text-generative ANN models such as GPT, unlike first-type ANN models, there is a divergence between input energy and information, which we associate with "predictions." These systems have an entirely different energy landscape, which is not represented by the energy of units, as in MLP models, but rather by the energy of connections between units, while the units themselves function only as static elements of the system's internal alphabet. For generative ANN models, these elements are, for example, words of a natural language (or their components), represented as tokens and their internal numerical representations - embeddings.

In general, static embeddings, such as those used in Word2Vec or GloVe, represent words as fixed-length vectors in a high-dimensional space. The proximity of these vectors (e.g., measured by cosine similarity) can serve as an indirect indicator of the similarity or association between words. This associative connection can be conditionally interpreted as a "probability of relation" between words, although this is not a direct probabilistic measure in the mathematical sense. Instead, it reflects only the averaged statistical relationships from the training corpus. Such a vector can be considered a representation of the "central meaning" of a word, but it does not account for polysemy or meaning shifts across different contexts.

In contrast, contextual embeddings in transformer-based models, such as BERT or GPT, dynamically transform a word's base representation—a lookup vector—based on surrounding words (context). This process highlights those aspects of the word's meaning that are relevant in the given context. The same word may have different vector representations depending on syntax, semantics, and even the text's style, allowing the model to account for polysemy. This mechanism enables the model to dynamically select—or rather compute—the most relevant (most probable, given a Softmax activation function) word relationships for a given context, based on the query-key-value attention matrix. The dynamics of changing relationships during inference are implemented through the Multi-Head Attention mechanisms across different layers of the model.

All word (embedding) relationships are formed based on weight coefficient matrices, which in turn are learned during training (in the energy interpretation—by "Maxwell's Demon"). This means that all potential pathways of word relationships are "embedded" in the model's weight matrices, as these matrices define the structure of possible associations and transformations. However, the specific relationships relevant to a given context are selected (computed) dynamically when processing input text. Thus, the model's weight matrices establish the depth of "potential channels" of relationships between units in the system's energy landscape in its resting state. The activation functions of the model's neurons determine the "output energy" from these channels, which in turn activates the "static energies" of the units. These unit energies collectively form the structured part of the system's output energy.

In this sense, the neurons of generative models effectively establish relationships between the elements of the system's internal alphabet (units). A simplified model of the energy landscape of a generative ANN is shown in Fig. 5.

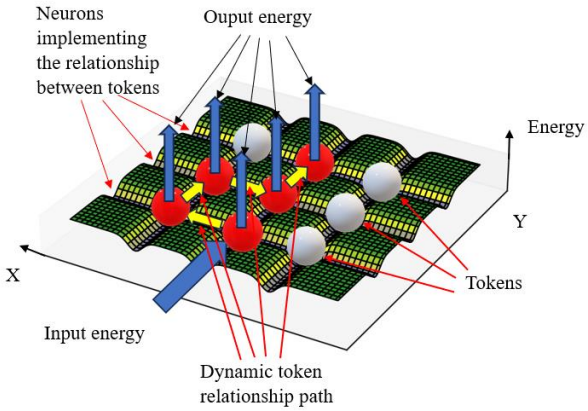

Fig. 5. The energy landscape model of generative ANN

Thus, this system implements the process of sequential perception of the energy (information) flow and its divergence across various "channels" of the system's energy landscape to activate the conditional energy of units (tokens).

Let us call such an energy model of an ANN a second-type model. This type may include recurrent ANNs as well as generative models based on transformers.

This model has a complex energy landscape formed by numerous weight matrices of various types (in transformer models, about 8 types of matrices are used for attention mechanisms, embeddings, linear transformations in fully connected layers, normalization, and output formation). It also consists of multiple hidden layers and neurons in each layer (for example, in the GPT-3 model, the hidden layer size can be up to 12,288 neurons for the largest model with 96 layers), which determines an enormous number of model parameters (GPT-3 contains up to 175 billion parameters). These parameters and their combinations form a vast space of possible microstates of the system.

Then, the dynamics of selecting the most relevant token (unit) connections during inference (the operation of a trained generative ANN) determine the reduction of the microstate space to its minimal possible extent (depending on the token selection mechanism). Ideally, it collapses to a single fixed set of microstates, which then defines the system's macrostate in the form of output information.

Thus, during training, the generative ANN model expands the space of possible system microstates (forms it), while during inference, it reduces this space to a minimum, which determines the system's macrostates. At the same time, a first-type model reduces the microstate space during training for a given macrostate.

It is evident that the information entropy at the output of a second-type ANN also decreases. This is because generative models select the most relevant connections during inference—that is, the most probable contextual structure of tokens. This process reduces the number of

possible token combinations, thereby decreasing the information entropy of the generated text, ideally collapsing it to a single structure for a given model.

However, in practice, heuristic algorithms such as Beam Search may be used to maximize the probability of a token sequence by selecting several of the most likely options. This leads to more predictable but less diverse results. This behavior can be seen as an analogy to the bottleneck effect in first-type ANNs.

To increase the diversity of texts (responses) in generative models (e.g., GPT-4), information entropy at the system's output is artificially increased. This is achieved by applying stochastic methods such as temperature scaling, which adds randomness to token selection, or Top-k/Top-p sampling, which limits the candidate pool for the next token selection, introducing additional randomness and enabling the generation of more diverse texts. Because of this, the space of microstates does not collapse into a single outcome but remains probabilistic, allowing for variation in model responses. Thus, the statistical nature of generative ANN training leads to the output of the most probable, i.e., contextually averaged, token structure. This naturally results in reduced information entropy at the system's output. However, as discussed earlier, the most informative and potentially valuable connection for problem-solving is the one with low probability, as it represents new or "creative" information.

We also see that reducing the number of possible system microstates during inference, leading to the formation of a single macrostate at the output, indicates a local reduction in the structured component of the system's thermodynamic entropy (26).

Thus, a second-type ANN model can be classified as an energy system with: *a deterministic static structure, a deterministic generalized energy landscape in a resting state, iteratively formed by "Maxwell's Demon", and a deterministic dynamic change in unit connection energy on the energy landscape in the process of minimizing information entropy and locally minimizing thermodynamic entropy*.

Conclusions. The results of these studies indicate the conceptual limitations of first-type and second-type ANN models. The structure of the internal energy of these systems in a resting state is formed by weight coefficients, which are obtained by another system—a "Maxwell's Demon"—that is external to the considered systems. These weight coefficients are formed based on statistical observations and are in no way connected to external energy (information). These limitations prevent the realization of a self-organization process in the system based on energy, as occurs in real physical systems, such as quantum systems.

Internal Energy of the Information System: A New Look

While information theory provides a quantitative description of uncertainty, and thermodynamics describes the distribution of energy, new approaches that characterize the energetic cost and the process of information processing for constructing an internal model and for the self-organization of the system remain an active area of fundamental and applied research in the physics of information, AI, and neuroscience [31–33].

At the core of the proposed theory is the concept of *internal energy within an information system and the construction of its energy landscape in an equilibrium state through the minimization of internal energy*. Modern ANNs, as informational models of biological systems, possess a conceptual limitation: all computational operations are performed using the symbols of an internal alphabet, represented by bits, bytes, and code tables — abstract substitutes for the system's external and internal energy. These alphabet symbols are static and lack the internal structure necessary to carry out the process of energy minimization.

In such systems, the training process requires the introduction of an additional conditional external energy – *pseudo-energy of mismatch* (error) between the system's current response and the desired response, which can be described by an energy function (analogous to a Lyapunov function) minimized via gradient descent along a pseudo-energy landscape. This pseudo-energy directly forms the dynamic structure of the system's internal energy, represented through the neural network's weight coefficients. The formation of pseudo-energy is based exclusively on statistical measurements, which – together with the static nature of the system's internal alphabet elements – frames our understanding of information from Shannon's perspective.

At the same time, biological systems directly convert external energy into internal energy within their sensory systems. This process is accompanied by the extraction of maximal information, not only through the detection of structural elements of external energy (analogous to the static elements of a system's internal alphabet), but also through the measurement of a multitude of their parameters. For example, the brain's visual system is capable of distinguishing not only linear segments but also their position, orientation, direction of motion, length, thickness, and spatial relations with other segments. All these parameters of external energy are transformed into the system's internal energy and represented as the reactions of interconnected groups of neurons – a pool or neuronal ensemble. This forms the initial representation of the internal energy structure of an energy landscape element — a unit corresponding to a single structural element of the system's static internal alphabet (e.g., a detected line segment).

The values obtained from these measurements can be projected onto internal quantitative scales and systems of coordinates/orientations. However, in order to form the internal energy landscape of the system and to perform gradient descent along this landscape during the minimization of internal energy, it is necessary to build a structure of these scales and coordinate/orientation systems, enabling transitions from high-energy quantitative parameter representations to lower-energy generalized or qualitative parameters during training. This can be achieved through the segmentation of scales and coordinate/orientation systems down to qualitative ordinal scales (such as "Greater-Equal-Less") or generalized coordinate/orientation systems (e.g., representing a rectangular coordinate system as half-planes). The process

of gradient descent formation along the energy landscape of a local structural element – a unit – during training is illustrated in Fig. 6.

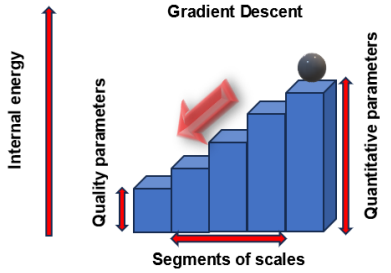

Fig. 6. The process of gradient descent along the energy landscape of a local structural element – a unit

The total internal energy of such a system, E_{total}, by analogy with a physical system, should be represented as the sum of the internal energies of its structural elements (units) and the energies of the interconnections between them:

$$E_{total} = \sum_i E_{unit,i} + \sum_{<j} E_{int,ij}, \qquad (35)$$

where E_{unit} is the internal energy of individual units, and E_{int} is the energy of the interconnections between units.

The energy of a unit is represented by a set of measurable external energy parameters k transformed by scaling functions h_k into internal energy values u_k on the corresponding measurement scales (coordinate systems). A *scaling function* is a function that "translates" a measurable (raw) parameter value into its "energy equivalent" on an internal scale. The set of scaling functions defines the multimodality of the system's energy landscape:

$$h_k: u_k \to \mathbb{R}_{\geq 0}; \qquad (36)$$

$$E_{unit,i} = \sum_k u_{k,i}. \qquad (37)$$

The process of gradient descent along the levels of a unit's internal energy during training can then be described by the following iterative relation:

$$u_{k,i}^{(L+1)} = u_{k,i}^{(L)} - \eta_k{}^{(L)} \frac{\partial E_{unit,i}^{(L)}}{\partial u_{k,i}^{(L)}}, \qquad (38)$$

where L is the energy landscape level during training (with the total internal energy of the system E_{total} decreasing at each subsequent level), and $\eta_k > 0$ is a parameter defining the step size for transitions between segments of the corresponding scale (coordinate system) or between scales with different levels of generality in the energy representation of the measured parameters. The set of these parameters and their selection procedures determine the *plasticity* of energy transitions, which may be associated with varying degrees of uncertainty.

The depth of the gradient descent along the energy levels of individual unit parameters depends on the frequency of their occurrence during training. This frequency is expressed through weight coefficients that directly determine the transitions between intervals, segments, or scales. The formation of the internal energy of interconnections between units is based on the comparison of values of similar parameters for units that possess a spatial-temporal relationship. In this case, a structural element of the connection energy c_{ij} between two units i and j is determined by the relation:

$$c_{k,ij} = h_c\left(\Phi\left(k_i, k_j\right)\right) \qquad (39)$$

where h_c is the corresponding scaling function used to compute the value of the interconnection energy $c_{k,ij}$, and Φ is a comparison function that defines the value of the interconnection parameter as the result of comparing the values of the same parameter k in both units. This equation thus links the energy of the units to the energy of their interconnections.

We can then define the gradient descent process along the energy levels of interconnections analogously to that of unit energies:

$$E_{int,ij} = \sum_j c_{k,ij}; \qquad (40)$$

$$c_{k,ij}^{(L+1)} = c_{k,ij}^{(L)} - \eta_c^{(L)} \frac{\partial E_{int,ij}^{(L)}}{\partial c_{k,ij}^{(L)}}. \qquad (41)$$

In this case, the depth of gradient descent along the energy levels of interconnections is determined by the minimal energy level of the common comparison scale. Additionally, the symmetry condition must be satisfied:

$$\left|c_{k,ij}\right| = \left|c_{k,ji}\right|. \qquad (42)$$

This process of gradient descent along the energy landscape of a local interconnection between two units is illustrated in Fig. 7.

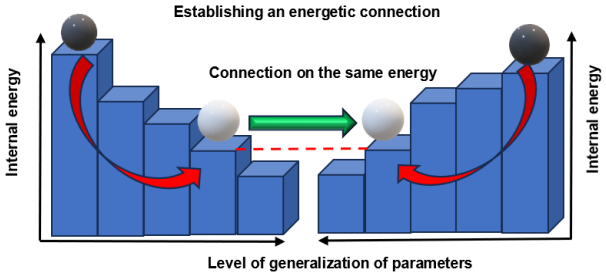

Fig. 7. The process of gradient descent along the energy landscape of the local relationship of two units

In general, the process of gradient descent along the system's energy landscape during training, in order to reach an equilibrium state at level L, is determined by the following conditions:

$$\frac{\partial E_{total}^{(L)}}{\partial u_i^{(L)}} = \frac{\partial E_{units}^{(L)}}{\partial u_i^{(L)}} + \frac{\partial E_{int}^{(L)}}{\partial u_i^{(L)}} = 0, \quad \forall i; \qquad (43)$$

$$\frac{\partial E_{total}^{(L)}}{\partial c_{ij}^{(L)}} = \frac{\partial E_{int}^{(L)}}{\partial c_{ij}^{(L)}} = 0, \quad \forall (i,j). \qquad (44)$$

At the same time, the interdependence

$$\frac{\partial^2 E_{int}^{(L)}}{\partial u_i^{(L)} \partial c_{ij}^{(L)}} \neq 0 \qquad (45)$$

guarantees that changes in the interconnection energy $c_{ij}^{(L)}$ affect the optimal distribution of $u_i^{(L)}$ and vice versa. These general conditions essentially demonstrate that the measurable/computable parameters belong to the same

segment, interval, or value of a scale/coordinate system, which characterizes their stability (invariance) to fluctuations in various samples of recognizable patterns.

Thus, the principal feature of implementing gradient descent along the system's energy landscape lies in the selection/determination of scales and coordinate/orientation systems for ordering the measurable or computable parameters of internal energy, as well as in harmonizing the segmentation thresholds for gradient transitions from specific parameters to their generalized representations. This is consistent with the findings of studies on neuronal and sensory structures in biological systems. For example, studies of receptive fields (RF) in the visual systems of humans and monkeys have shown that their segmentation and overlap form a "grid" of spatial coordinates – an internal coordinate and orientation system [34].

Investigations indicate that constructing the energy landscape of an informational system is associated with identifying stable substructures, separated by critical points that define the disruption of stable parameters of interconnections between structural elements (units) within these substructures. These critical points serve as important components shaping the structure of an energy attractor — the concept of a holistic image [35].

One of the conceptual propositions of these studies is the formalization of the definition of information in terms of energy. In the article [35] information is not considered abstractly but is explicitly defined as *subjectively measurable parameters of external energy structures*. The processes of self-organization of a system's energy structure and its minimization during evolution (training) are based on the application of proposed reduction operators, presented as energy functions (Lyapunov functions), without invoking a "Maxwell's Demon". These operators include: the parametric reduction operator $R_{u,c}$, the structural reduction operator R_w, and the structural-parametric reduction operator R_{sp}. A generalized reduction operator was also introduced in the study, based on the composition of these specific operators:

$$E_{red} = R_w(R_{sp}(R_{u,c}(E_{total}^{(0)}))) < E_{total}^{(0)}, \qquad (46)$$

where $E_{total}^{(0)}$ is the total energy of the system at the initial structural level, and E_{red} is the minimized (reduced) internal energy.

It was formally demonstrated that the application of these operators leads the system to converge toward stable states with minimal internal energy – attractors. These attractors (*Con*) represent stable (invariant) internal structures corresponding to recognized patterns – internally formed models of the external world *M*.

$$M = R(S_{int}(P(E_{int})) = Con; \qquad (47)$$

$$P(E_{int}) = D\big(S_{ext}(E_{ext})\big), \qquad (48)$$

where *R* is the generalized reduction operator, $S_{int}(P(E_{int}))$ is the structure of the set of parameters *P* of the system's internal energy E_{int}, and *D* is the generalized detection (measurement) operator for structured S_{ext} (having spatiotemporal structure) parameters of external energy E_{ext}.

It was also shown that the model *M* of an individual holistic image can be represented as a *hypergraph, and the evolution of M as a process of its reduction (minimization)*. The study [35] also proposes and examines a hypothesis that information can be regarded as *the primary factor in the phase transition of the structured part of external energy (Gibbs free energy formed by the external system) into the structure of internal energy, which is considered as a model of the external world*.

Experimental verification of the theory

As part of the experimental studies, a model was proposed consisting of an interconnected structure of image-detector neurons (local attractors) named the Compartment Attractor Neural Network (ComAN), which represents a further development of the model introduced by Yuri Parzhin, Viktor Kosenko et al. (2020) [36]. The choice of this name is due to the fact that in the proposed model, each independent detector neuron simulates the information-energy processes occurring in the dendritic tree of a biological neuron and memorizes an attractor structure based on internal energy. This name emphasizes that the detector neurons possess an internal compartmental structure that helps to form stable states (attractors) for the detection of holistic images. This approach is inspired by the neurobiological paradigm of concept neurons [37–39].

To empirically confirm the key principles of the proposed theory, a concept for a computational experiment using ComAN was developed. ***The aim of this experiment*** *is to demonstrate that a system based on the principles of internal energy minimization can self-organize and evolve through direct learning for pattern recognition, without relying on traditional externally controlled learning algorithms based solely on statistical data.*

The goal of the experiment is achieved by demonstrating the high generalization ability of the model in the process of one-pass training without using the backpropagation algorithm on ultra-small training data sets.

One of the main challenges in conducting such an experiment lies in implementing the stage of preliminary information processing that simulates the operation of a complex system of visual receptors and pre-detector neurons of holistic images in the visual cortex. These include, for example, neurons responding to the orientation of line segments or curve features, the length or direction of motion of segments, to segment ends or angular points, as well as to more complex and specific features of individual structural elements and holistic images [40-42]. According to the proposed theoretical concept, this modeling involves selecting the types of detectable structural elements within input images and their measurable parameters, as well as mapping these values onto preselected and segmented quantitative and qualitative measurement scales and coordinate systems.

To simplify these tasks and focus on the fundamental principles of self-organization, two-dimensional images of handwritten digits from the classic benchmark MNIST database were selected for classification. The choice of this library also simplified the task of scaling images. The structural elements detected at the preprocessing stage were defined as: line segments, endpoints (segment ends), angular points, points of intersection and junctions of segments, and whether the contour of the image was closed or open. Additionally, primary measurable parameters of the structural elements were established, including: the

orientation of line segments in a segmented relative coordinate system, their position in an absolute coordinate system (receptive field), the length of the segments, and the positions of structural points in both absolute and relative coordinate systems.

This selection made it possible to determine the types and values of secondary parameters, which define the relationships between structural elements based on the comparison of their primary parameters. The values of these secondary parameters are also mapped (projected) onto segmented quantitative and qualitative scales and coordinate/orientation systems of the segments. All these parameters and their interrelationships form the basis of the internal energy landscape of the system. As demonstrated in the theoretical part of the study, the segmented structure of quantitative and qualitative measurement scales and coordinate/orientation systems forms the foundation of the gradient descent process along the system's energy landscape during training.

This approach required representing images of handwritten digits in the form of skeletonized contours approximated by straight line segments.

To model the system's energy landscape, hypergraphs were used. For working with hypergraphs, the Python **NetworkX** library and the **Neo4j** graph database were selected. This choice was driven by the need to work with large sets of structured data. For conducting the computational experiment, a general scheme (pipeline) for the experimental verification of the proposed concept was developed, as presented in Fig. 8.

This scheme includes the following main blocks:

1. *Data Sources*. The system uses two datasets: a training set for constructing attractors and a test set for evaluating classification accuracy.

2. *Connector (Preprocessing)*. Incoming requests are processed by the connector module, which standardizes the data format.

3. *Skeletonization Pipeline*. This stage involves several steps: binarizing the MNIST images; skeletonizing the images to obtain a contour composed of line segments while preserving the topological structure of the original image; applying the Growing Neural Gas (GNG) algorithm to convert the resulting skeleton into a graph; and using the Ramer–Douglas–Peucker (RDP) approximation algorithm to simplify the resulting graph by reducing the number of angular points (and segments) during the curve approximation with line segments. The resulting graph is serialized using NetworkX into JSON format and published to a Kafka topic.

4. *Contour Analysis*. At this stage, various analyzers are used: analyzers based on contour traversal for sequential feature (parameter) extraction, and global analyzers that evaluate the entire hypergraph structure to extract holistic features such as exposure analysis.

5. *Concept (Energy Attractor) Formation During Training*. At the training stage, the system constructs concepts – generalized hypergraphs – using: identification of critical points that define key nodes in the hypergraph; searching for the maximum common sub-hypergraph between two compared hypergraphs (the modifiable and a new training example); reduction of the modifiable hypergraph – structural elements, connections, and substructures – based on the application of the developed energy reduction operators; and iterative generalization of the attractor hypergraph.

6. *Sample Classification During Inference*. At this stage, the system: reduces the hypergraph of the classified sample using reduction operators; compares the reduced hypergraph with all existing concepts to determine their structural similarity (isomorphism); activates class detector neurons with concepts that fully match the "incoming" hypergraph or sub-hypergraph; and based on a "Winner Take All" competition mechanism among simultaneously excited detector neurons, determines the classification result produced by the ComAN network.

Thus, the model uses one neuron for each data class. Each neuron simulates information processing by the dendritic tree of a biological neuron (Poirazi & Mel, 2001 [43]) in the form of constructing and modifying a hypergraph during direct learning (without backpropagation of error) until a stable structure — an attractor — is achieved, which corresponds to the minimum of the system's internal energy. The process of minimizing internal energy is governed by reduction operators, which are represented as Lyapunov functions.

In preliminary experiments, the viability of the proposed theoretical concept was confirmed based on intermediate results, which demonstrated classification accuracy for six structurally different classes of handwritten digits: "1", "2", "3", "6", "7", and "9" according to various metrics: **Accuracy: 82.44%,** Precision: 83.33%, Recall:

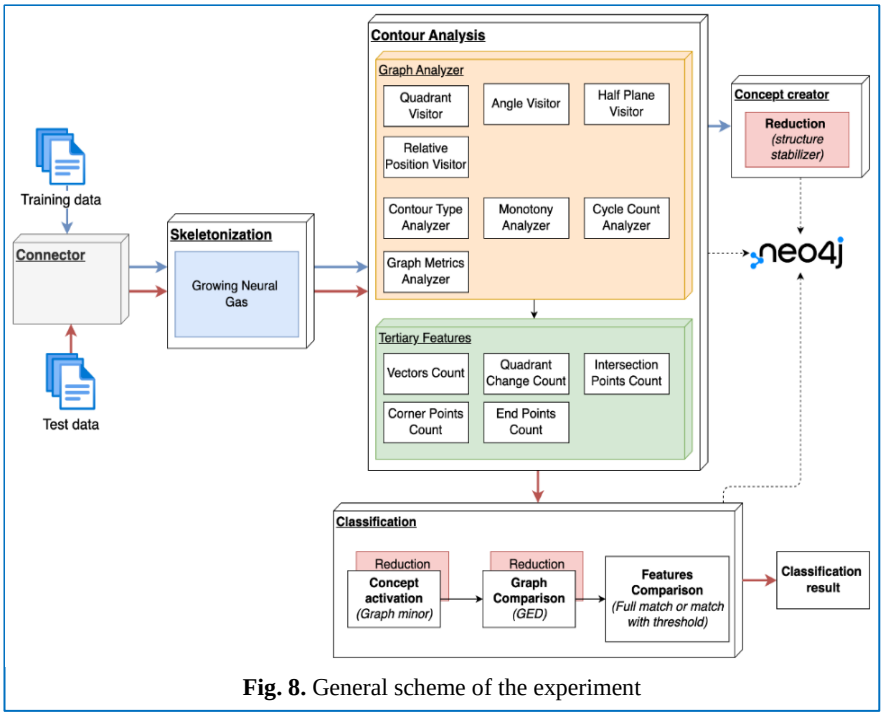

Fig. 8. General scheme of the experiment

82.44%, and F1 Score: 82.25%. These classes were selected to balance simple and complex structures in terms of attractor construction. The number of classes will be increased as the model evolves.

Each class neuron is trained using only **5–6 unique examples** (up to 36 examples for the 6 selected classes). For each example, augmentation is applied — up to 10 additional augmented examples. The total training dataset contains only **350 examples**. To test the model, a test dataset consisting of **4,734** previously unseen MNIST examples is used. The model is not prone to overfitting and is invariant to changes in image scale.

Although this classification problem formulation fundamentally differs from classification tasks based on statistical learning approaches, we can nevertheless compare the preliminary testing results of the proposed model with known models operating under similar conditions with extremely small training datasets — Few-shot Learning (FSL) models. The comparison results are presented in Table 1.

Table 1. **Comparison of simulation results**

Model	Minimum number of unique samples	Augmentation	Epochs	Accuracy (%)	Source	Notes
ComAN Model (6 classes)	Up to 36 (5–6/class)	Yes (up to 10 per class – 350 total)	1	82.44	This work	No backpropagation, no hyperparameter tuning
SVM (RBF) Reduced MNIST (RMNIST/5-10)	50–100 (5-10/class)	-	1	69–75	Nielsen [44]	Requires hyperparameter tuning
MLP	200 (20/class)	No/Yes	100	53/61	Zhang [45]	Backpropagation, hyperpara-meter tuning, overfitting problem
CNN (FMNIST)	100–200 (10-20/class)	Yes	50	74–78	Brigato et al. [46]	Backpropagation, hyperparameter tuning
Prototypical Networks (ProtoNet)	5/class (5-way) miniImageNet	-	FSL (Metrics training)	71	Chen et al. [47], Snell et al. [48]	Pre-training in basic classes, using 5-way classification
Model-Agnostic Meta-Learning (MAML)	5/class (5-way) miniImageNet	-	FSL (Meta-learning)	72	Chen et al. [47]	Pre-training on task distribution, learning initialization parameters for fast adaptation
CNN (RMNIST/5)	50 (5/class)	Yes (500–1000)	50	84.38	Nielsen [44]	Backpropagation, dropout, hyperparameter tuning

Conclusions

The obtained results demonstrate the unique advantages of the developed model:

- **No backpropagation.** Unlike most models, this system employs energy attractors, minimizing internal energy, bringing it closer to biologically inspired models.
- **One-pass training in a single epoch on small datasets** is a unique feature that allows abandoning the concept of statistical learning in favor of biologically inspired Hebbian frequency-based learning.
- **Energy dynamics of attractors.** The implementation of model dynamics based on Lyapunov functions aligns it with physical and biological systems. This model represents virtually the first working implementation of a dendritic hypergraph model with internal energy minimization via Lyapunov functions on the MNIST dataset.
- **Biological plausibility of the model** is confirmed by direct modeling of dendritic processing, which is absent in classical deep learning models. The model demonstrates a neuromorphic architecture close to that of a biological neuron.
- **Achieved testing results** — 82.44% accuracy on 4,734 previously unseen samples after one-pass training on only 5–6 unique examples per class – indicate extremely efficient learning and a very high level of generalization of key classification features under severely limited training data.
- **Scalability:** the model can be easily scaled to any number of classes without the need to retrain already existing neurons.
- **No overfitting:** the model is not prone to overfitting, as a stable attractor is formed for each individual class during training.

At the same time, these results reveal certain limitations and drawbacks of the current stage of modeling:

- The simplified sensory preprocessing model currently limits the achievable accuracy ceiling and the applicability of the model to the recognition of more complex and dynamic patterns.
- Dependence on custom tools based on Neo4J and NetworkX also constrains the scalability and performance of the model.
- The absence of evolutionary biological inter-neuronal connection modeling leads to significantly increased computational costs when analyzing hypergraphs.

We believe that the impact of these limitations on the model's efficiency will be substantially reduced in the course of further research.

REFERENCE

1. Liu, Y., Cao, J., Liu, C., Ding, K. and Jin, L. (2024), "Datasets for Large Language Models: A Comprehensive Survey", *arXiv*, arXiv:2402.18041, doi: https://doi.org/10.48550/arXiv.2402.18041

2. Villalobos, P., Ho, A., Sevilla, J., Besiroglu, T., Heim, L. and Hobbhahn, M. (2024), "Will we run out of data? Limits of LLM scaling based on human-generated data", *arXiv*, arXiv:2211.04325v2, available at:: https://arxiv.org/html/2211.04325v2
3. Henshall, W. (2024), "The Billion-Dollar Price Tag of Building AI", *TIME*, available at: https://time.com/6984292/cost-artificial-intelligence-compute-epoch-report/?utm_source=chatgpt.com
4. Zewe, A. (2025), "Explained: Generative AI's environmental impact", *MIT News*, Massachusetts Institute of Technology, January 17, available at: https://news.mit.edu/2025/explained-generative-ai-environmental-impact-0117
5. Luccioni, S., Trevelin, B. and Mitchell, M. (2024), "The Environmental Impacts of AI – Primer", *Hugging Face*, published September 3, available at: https://huggingface.co/blog/sasha/ai-environment-primer?utm_source=chatgpt.com
6. Zhou, L., Schellaert, W., Martínez-Plumed, F., Moros-Daval Y., Ferri C. and Hernández-Orallo J. (2024), "Larger and more instructable language models become less reliable", *Nature*, vol. 634, pp. 61–68, doi: https://doi.org/10.1038/s41586-024-07930-y
7. Williams, R. (2024), "Why Google's AI Overviews gets things wrong", *MIT Technology Review*, doi: https://www.technologyreview.com/2024/05/31/1093019/why-are-googles-ai-overviews-results-so-bad/
8. LeCun, Y. (2022), "A Path Towards Autonomous Machine Intelligence", *OpenReview*. 62 p., available at: https://openreview.net/pdf?id=BZ5a1r-kVsf
9. Marcus, G. (2018), "Deep Learning: A Critical Appraisal", *arXiv*, *arXiv.1801*.00631, doi: https://doi.org/10.48550/arXiv.1801.00631
10. Zador, A. (2019), "A critique of pure learning and what artificial neural networks can learn from animal brains", *Nat. Commun.*, vol. 10, 3770, doi: https://doi.org/10.1038/s41467-019-11786-6
11. Shumailov, I., Shumaylov, Z., Zhao, Y., Papernot, N., Anderson, R. and Gal, Ya. (2024), "AI models collapse when trained on recursively generated data", *Nature*, vol. 631, pp. 755–759, doi: https://doi.org/10.1038/s41586-024-07566-y
12. Alemohammad, S., Casco-Rodriguez, J., Luzi, L., Humayun, A.I., Babaei, H., LeJeune, D., Siahkoohi, A. and Baraniuk, R.G. (2023), "Self-Consuming Generative Models Go MAD", *arXiv*, arXiv:2307.01850, doi: https://doi.org/10.48550/arXiv.2307.01850
13. Glansdorff, P. and Prigogine, I. (1971), *Thermodynamic Theory of Structure, Stability and Fluctuations*, Wiley-Interscience, 306 p., available at: https://archive.org/details/thermodynamicthe0000glan/page/n5/mode/2up
14. Hopfield, J. (1982), "Neural networks and physical systems with emergent collective computational abilities", *Proc. Natl. Acad. Sci. U. S. A.* vol. 79(8), pp. 2554–2558, available at:: https://www.pnas.org/doi/epdf/10.1073/pnas.79.8.2554
15. Hopfield, J. (1984), "Neurons with graded response have collective computational properties like those of two-state neurons", *Proc. Natl. Acad. Sci. U.S.A.* vol. 81, pp. 3088–3092, available at:: https://www.pnas.org/doi/10.1073/pnas.81.10.3088
16. Krauth, W. (2026), *Statistical Mechanics: Algorithms and Computations*, Oxford University Press, 360 p., available at: https://global.oup.com/booksites/content/9780198515364/
17. Sassano, M. and Astolfi, A. (2013), "Dynamic Lyapunov functions", *Automatica*, vol. 49. pp. 1058–1067, available at: https://liberzon.csl.illinois.edu/teaching/dynamic-Lyapunov-functions.pdf?utm_source=chatgpt.com
18. Walter J. and Barkema, G. (2015), "An introduction to Monte Carlo methods", *Physica A: Statistical Mechanics and its Applications*, vol. 418, pp. 78–87, doi: https://doi.org/10.48550/arXiv.1404.0209
19. LeCun, Y., Chopra, S., Hadsell, R., Ranzato M. and Huang, F. (2006), "A Tutorial on Energy-Based Learning", *MIT Press*, available at: http://web.stanford.edu/class/cs379c/archive/2012/suggested_reading_list/documents/LeCunetal06.pdf
20. Belanger, D. and McCallum, A. (2016), "Structured Prediction Energy Networks", *Proceedings of The 33rd Int. Conference on Machine Learning*, *PMLR*, vol. 48, pp. 983–992, available at: https://proceedings.mlr.press/v48/belanger16.html
21. Grathwohl, W., Wang, K., Jacobsen, J., Duvenaud, D., Norouzi, M. and Swersky K. (2020), "Your Classifier is Secretly an Energy Based Model and You Should Treat it Like One", *arXiv*, doi: https://doi.org/10.48550/arXiv.1912.03263
22. Friston, K. (2012), "A Free Energy Principle for Biological Systems", *Entropy (Basel)*, vol. 14(11), pp. 2100–2121, available at: https://pubmed.ncbi.nlm.nih.gov/23204829/
23. Bengio, Y., Lee, D., Bornschein, J., Mesnard, T. and Lin, Z. (2016), "Towards Biologically Plausible Deep Learning", *arXiv*, arXiv.1502.04156, doi: https://doi.org/10.48550/arXiv.1502.04156
24. Landauer, R. (1961), "Irreversibility and Heat Generation in the Computing Process," in *IBM Journal of Research and Development*, vol. 5, no. 3, pp. 183-191, doi: https://doi.org/10.1147/rd.53.0183
25. Leff, H. and Rex, A. (1990), *Maxwell's Demon: Entropy, Information, Computing (Princeton Series in Physics)*, Princeton University Press. 349 p., ., available at: https://www.amazon.co.uk/Maxwells-Demon-Entropy-Information-Computing/dp/0750300574
26. Middelburg, J. (2024). "The Gibbs Free Energy", *Thermodynamics and Equilibria in Earth System Sciences: An Introduction. SpringerBriefs in Earth System Sciences. Springer*, *Cham*, pp. 35–46, doi: https://doi.org/10.1007/978-3-031-53407-2_4
27. Bishop, C. (2006), *Pattern Recognition and Machine Learning*. Springer, 778 p., available at: https://link.springer.com/book/
28. Brown, E. and Ahlers, G. (2008), "A model of diffusion in a potential well for the dynamics of the large-scale circulation in turbulent Rayleigh-Benard convection", *arXiv*, 0807.3193, 43 p., available at: https://arxiv.org/pdf/0807.3193
29. Shwartz-Ziv, R. and Tishby, N. (2017) "Opening the Black Box of Deep Neural Networks via Information", *arXiv*, arXiv.1703.00810, doi: https://doi.org/10.48550/arXiv.1703.00810
30. Tishby, N., Pereira F. and Bialek, W. (2000), "The Information Bottleneck Method", *arXiv*, arXiv:physics/0004057, available at: https://arxiv.org/abs/physics/0004057
31. Dawid, A. and LeCun, Y. (2023), "Introduction to Latent Variable Energy-Based Models: A Path Towards Autonomous Machine Intelligence", *arXiv*, arXiv:2306.02572, doi: https://doi.org/10.48550/arXiv.2306.02572
32. Liu, J., Meng, Y., Fitzsimmons, M. and Zhou, R. (2025), "Physics-informed neural network Lyapunov functions: PDE characterization, learning, and verification", *Automatica*, vol. 175, pp. 112–193, doi: https://doi.org/10.1016/j.automatica.2025.112193
33. Acharya, J., Basu, A., Legenstein, R., Limbacher, T., Poirazi, P. and Wu, X. (2022), "Dendritic Computing: Branching Deeper into Machine Learning", *Neuroscience*. vol. 489, pp. 275–289, doi: https://doi.org/10.1016/j.neuroscience.2021.10.001
34. Lerma-Usabiaga, G., Winawer, J. and Wandell, B. (2021), "Population Receptive Field Shapes in Early Visual Cortex Are Nearly Circular", *Journal of Neuroscience*, vol. 41(11), pp. 2420–2427; doi: https://doi.org/10.1523/JNEUROSCI.3052-20.2021
35. Parzhyn, Y. (2025), "Architecture of Information", *arXiv*, arXiv:2503.21794, doi: https://doi.org/10.48550/arXiv.2503.21794
36. Parzhin, Y., Kosenko, V., Podorozhniak, A., Malyeyeva, O. and Timofeyev, V. (2020), "Detector neural network vs connectionist ANNs", *Neurocomputing*, vol. 414, pp. 191–203, doi: https://doi.org/10.1016/j.neucom.2020.07.025
37. Quiroga, R., Reddy, L., Kreiman, G. Koch, C. and Fried, I. (2025), "Invariant visual representation by single neurons in the human brain", *Nature*, vol. 435, pp. 1102–1107, doi: https://doi.org/10.1038/nature03687

38. Bowers, J. (2009), “On the Biological Plausibility of Grandmother Cells: Implications for Neural Network Theories in Psychology and Neuroscience”, *Psychological review*, vol. 116, pp. 220–251, doi: https://pubmed.ncbi.nlm.nih.gov/19159155/
39. Roy, A. (2013), “An extension of the localist representation theory: grandmother cells are also widely used in the brain”, *Front. Psychol.*, vol. 4, doi: https://doi.org/10.3389/fpsyg.2013.00300
40. Gawne, T. (2015), “The responses of V1 cortical neurons to flashed presentations of orthogonal single lines and edges”, *J Neurophysiol*, vol. 113, is. 7, pp. 2676–2681, doi: https://doi.org/10.1152/jn.00940.2014
41. Shevelev, I. (1998), “Second-order features extraction in the cat visual cortex: selective and invariant sensitivity of neurons to the shape and orientation of crosses and corners”, *Biosystems*, vol. 48, pp. 195–204, doi. https://doi.org/10.1016/s0303-2647(98)00066-5
42. Leopold, D., Bondar, I. and Giese, M. (2006), “Norm-based face encoding by single neurons in the monkey inferotemporal cortex”, *Nature*, vol. 442, pp. 572–575, doi: https://doi.org/10.1038/nature04951
43. Poirazi, P. and Mel, B. (2021), “Impact of active dendrites and structural plasticity on the memory capacity of neural tissue”, *Neuron*, vol. 29, pp. 779–796, doi: https://doi.org/10.1016/s0896-6273(01)00252-5
44. Nielsen, M. (2017), “Reduced MNIST: how well can machines learn from small data?”, *Cognitive Medium*, available at: https://cognitivemedium.com/rmnist?utm_source=chatgpt.com
45. Zhang, L. (2019), *Overfitting problem*, University of Toronto, available at: https://www.cs.toronto.edu/~lczhang/360/lec/w05/overfit.html
46. Brigato, L., and Iocchi, L. (2020), "A Close Look at Deep Learning with Small Data", *arXiv*, arXiv:2003.12843, doi: https://doi.org/10.48550/arXiv.2003.12843
47. Chen, W., Liu, Y., Kira, Z., Wang, Y. and Huang, J. (2019), A closer look at few-shot classification”, *ICLR 2019*, 16 p., available at: https://openreview.net/pdf?id=HkxLXnAcFQ
48. Snell, J., Swersky, K. and Zemel, R. (2017), “Prototypical Networks for Few-shot Learning”, *arXiv*, arXiv:1703.05175, doi: https://doi.org/10.48550/arXiv.1703.05175

Received (Надійшла) 28.05.2025
Acceptedforpublication (Прийнята до друку) 13.08.2025

ВІДОМОСТІ ПРО АВТОРІВ / ABOUT THE AUTHORS

Паржин Юрий Володимирович – доктор технічних наук, пост докторант Школи Комп'ютерних та Кібер Наук Університету Огасти, Огаста, Сполучені Штати Америки;
Yurii Parzhyn – Doctor of Technical Sciences, Postdoctoral Fellow, School of Computer and Cyber Sciences, Augusta University, Augusta, United States of America.
e-mail: yparzhyn@augusta.edu; ORCID Author ID: https://orcid.org/0000-0001-5727-1918;
Scopus ID: https://www.scopus.com/authid/detail.uri?authorId=57224412390.

Лапін Микита Олексійович – аспірант кафедри системного аналізу та інформаційно-аналітичних технологій, Національний технічний університет "Харківський політехнічний інституту", Харків, Україна;
Mykyta Lapin – Postgraduate student of the Department of System Analysis and Information and Analytical Technologies, National Technical University "Kharkiv Polytechnic Institute", Kharkiv, Ukraine.
e-mail: Mykyta.Lapin@cit.khpi.edu.ua; ORCID Author ID: https://orcid.org/0009-0003-6307-1172.

Бохан Костянтин Олександрович – кандидат технічних наук, доцент кафедри системного аналізу та інформаційно-аналітичних технологій, Національний технічний університет "Харківський політехнічний інституту", Харків, Україна;
Kostiantyn Bokhan – PhD, Associate Professor of the Department of System Analysis and Information and Analytical Technologies, National Technical University "Kharkiv Polytechnic Institute", Kharkiv, Ukraine.
e-mail: kostiantyn.bokhan@khpi.edu.ua; ORCID Author ID: https://orcid.org/0000-0003-3375-2527.
Scopus ID: https://www.scopus.com/authid/detail.uri?authorId=57191592568.

Новий підхід до побудови енергетичних моделей нейронних мереж

Ю. В. Паржин, М. О. Лапін, К. О. Бохан

Анотація. Актуальність. Сучасні моделі штучних нейронних мереж потребують багато енергії та інших витрат на їх навчання та функціонування. Для навчанні генеративних моделей використовуються величезні обсяги даних. В той же час, такі моделі мають проблеми з довірою щодо згенерованої ними інформації. Альтернативою сучасним парадигмам побудови та навчання нейронних мереж є розробка енергетичних моделей, які потенційно повинні позбутися цих недоліків та наблизити процес обробки інформації до біологічно та фізично обґрунтованого процесу. Але існуючи енергетичні моделі мало чим відрізняються від класичних моделей з точки зору їх недоліків та обмежень. Тому розробка нових підходів в моделюванні енергетичних процесів обробки інформації в нейронних мережах є актуальною. **Об’єкт дослідження** – процес обробки інформації в штучних нейронних мережах. **Предмет дослідження** – математичні моделі побудови та навчання штучних нейронних мереж. **Метою даної статті** є розроблення та експериментальна перевірка теоретичної бази, що постулює енергетичну природу інформації та її роль у самоорганізації та еволюції складних інформаційних систем. **Результати дослідження.** Запропоновано фундаментальну теорію, що описує інформацію як структуру сприйманих параметрів зовнішньої енергії, яка керує процесами формування внутрішньої енергетичної структури системи – її моделі зовнішнього світу. Ця теорія включає концепції енергетичних ландшафтів, принципи структурної та параметричної редукції, заснованої на енергії, а також критичний аналіз існуючих обчислювальних парадигм. Проведені експериментальні дослідження з побудови та навчання розробленої енергетичної моделі підтверджують її високу узагальнюючу здатність в процесі однопрохідного навчання без використання алгоритму зворотного поширення помилки на надмалих навчальних наборах даних.

Ключові слова: енергозберігаючі обчислення; енергетичний ландшафт; штучні нейронні мережі; нейрообчислення; інформація; ентропія.