ISSN: 2834-7706

Research Article

Journal of Mathematical Techniques and Computational Mathematics

Will Artificial Intelligence Become Conscious? Can Thermodynamics Explain the Evolution of Intellect?

Eva Deli

University of Debrecen, Department of Anatomy, Histology and Embryology.

***Corresponding author:**
Eva Deli, University of Debrecen, Department of Anatomy, Histology and Embryology 3420 East Anderson drive, Phoenix, AZ 85032, United States.

Submitted: 15 Oct 2022; **Accepted:** 27 Oct 2022; **Published:** 16 Nov 2022

Citation: Deli, E. (2022). Will Artificial Intelligence Become Conscious? Can Thermodynamics Explain the Evolution of Intellect? *J Math Techniques Comput Math*, 1(2), 124-128.

Abstract

Deep neural networks (DNNs), founded on the brain's neuronal organization, can extract higher-level features from raw input. However, complex intellect via autonomous decision-making is way beyond current AI design. Here we propose an autonomous AI inspired by the thermodynamic cycle of sensory perception, operating between two information density reservoirs. Stimulus unbalances the high entropy resting-state and triggers a thermodynamic cycle. By recovering the initial conditions, self-regulation generates a response while accumulating an orthogonal, holographic potential. The resulting high-density manifold is a stable memory and experience field, which increases future freedom of action via intelligent decision-making.

Key Words: Deep Neural Networks, Artificial Intelligence, Intelligent Computation, Orthogonal Transformation, Entropy, Carnot Cycle.

Introduction

Although homeostasis, i.e., a dynamic equilibrium is the fundamental characteristic of all life, survival requires constant interaction with the environment. The processing of stimuli by the sensory system generates memory, conserved in the synaptic map. Therefore, the synaptic map represents an abstract representation of the environment permitting purposeful behavior to even the most primitive animals. Meaning generation, a fundamental character of intellect, lends a predictive ability for cognition [1, 2]. The brain's sensory cycle revolves between an information-rich source (the environment) and an information-hungry sink (the brain), forming a thermodynamic cycle. Recently, the Carnot engine was used to analyze the brain's information processing [3-5].

Deep learning (DL) is an AI function that mimics the human brain's processing of data. Recurrent neural networks (RNNs) are characterized by the presence of feedback connections in a hidden layer, which allows generating a state–space representation that equips the network with short-term memory capability. RNNs are universal approximators of dynamical systems, meaning that, given enough neurons in the hidden layer, it is possible to fine-tune the weights to achieve any desired level of accuracy.

As the AI system moves through hidden layers, it manipulates the data by extracting higher-level features from raw input [6]. Despite their great success, there is still no comprehensive understanding of the optimization process or the internal organization of DNNs, and they are often criticized for being used as mysterious "black boxes". A "black box" lacks transparency of how input data are transformed to model outputs. In the following, we analyze the DNN's "black box" to formulate a deeper understanding [7].

Discussion

Holographic Considerations (The Field Representation of Complexity)

The second law of thermodynamics is a fundamental law of physics. Entropy maximization might be a general phenomenon in physics, cosmology, computer science, and animal behavior [8-10]. Recent works also associated entropy maximization with the intellectual ability to increase future freedom of action [8, 11, 12].

Information consumption embeds the brain within the environment's energy-information cycle. Sensory perception is an automatic and involuntary process. For example, reading road signs is instinctive because sensory stimuli push the brain's energy balance out of equilibrium.

Response to a stimulus depends on an intimate understanding of the environment. Information compression generates an orthogonal transformation [13, 14, 15]. Thus, the constantly updating synaptic map accumulates memory (Figure 1) via a holographic, high-fidelity manifold [16]. Orthogonal transformation produces

a temporal orientation with a predictive nature [2, 17]. Nevertheless, the brain's resting or task-negative state, occurring without explicit task, forms the self, which remains stable throughout life Therefore, engagement with the physical environment projects the physical laws of the environment into a temporal system, the mind. Thus, the mind is 'about' the world by being foremost about itself.

Similar to the human mind, deep neural networks (DNNs) accumulate intellect from experience. Like in the human brain, each node of the neural network is responsible for solving a small part of the problem. Experience accumulates by being passed further from neurons to other neurons, until the interconnected nodes are able to solve the problem and give an output. Trial and error are key to the nodes ability to learn.

In the computer brain, like in the organic brain, the intelligent response depends on continuously changing energy balance toward the equilibrium. Transmission strengthens connections between units, giving rise to segregated, hierarchical, and modular structures. In the following, we analyze artificial networks from thermodynamic principles [4, 5, 18, 19].

Thermodynamic considerations

Physical processes can be dissipative, which reconstructs the past, and intelligent, those that anticipate the future [20]. The first kind, exothermic process dumps entropy, and energy into its environment, whereas the latter, endothermic actions absorb entropy while requiring energy to operate. Exothermic processes make endothermic ones possible [20]. Intelligent computation belongs to the second category. These endothermic processes control the future by boosting mental abilities [3, 8, 21]. Thus, intellect is the evolution of the synaptic map through the thermodynamic regulation of the cortical brain.

Information science interprets entropy as the "amount of information" contained in a variable. The sensory system compresses spatial stimulus into a holographic representation. Data compression increases information density [22, 23, 24]. The orthogonal transformation integrates temporally distant identities into a subjective observer state [1, 2, 11, 15, 19]. Therefore, while temporal variability affords high degrees of freedom, the transformation of sensory data by an orthogonal morphing generates stability for experience and memory, independent from sensory disturbance [15]. Perception (response) results from self-regulation, restoring the high entropy (resting) state. Therefore, resting homeostasis is an entropic requirement. Because the evoked states build on the resting potential, they can be analyzed using thermodynamics [3, 4, 5, 18].

In biological brains sleep modifies synaptic weights according to spike-timing-dependent plasticity rules [25]. Recently, in neural networks, an offline sleep-like phase modified the connection strength [26]. Introducing sleep into artificial neural networks could counter catastrophic forgetting by recovering forgotten tasks. Similarly, resting dynamics strengthen basic network features. Furthermore, high entropy oscillations' ability to access microstates (generating configurations with equal likelihood) increases the degrees of freedom or the freedom of thought.

Figure 1: Mental field evolution the thermodynamic cycle (indicated by ovals) represents a fluid, chaotic and reversible process. The energy field (solid line) represents the increasing complexity of the neural system.

Intelligent Computation via the Reversed Carnot Cycle

The Carnot cycle is an idealized theoretical process between two heat reservoirs with vanishing net entropy production. As the theoretical framework of the reversed Carnot cycle explains the brain evoked cycle, it might also account for the "black box" of the DNN (Figure 2). The reversed Carnot cycle absorbs heat (information) from a low-temperature reservoir (environment) and delivers it to a hot reservoir (the internal organization of the DNN), which requires work input. The product $\Delta T \Delta H$ (temperature and entropy) is the cycle's area and the energy E required to complete a processing cycle. The efficiency is determined by the average connection strength between the states; therefore, learning takes many cycles.

Information is physical. Landauer has shown that the transformation between energy and information has energetic consequences (Landauer, 1961; Landauer, 1991) [27, 28]. It requires energy to erase information from the computing device; therefore, erasing a bit of information releases a minimum heat. Landauer's principle is the ejection of (all or part of) some correlated information from a controlled, digital form (e.g., a computed bit) to an uncontrolled, non-computational form, i.e., as part of a thermal environment. An irreversible, permanent increase in entropy of $\log 2 = k \ln 2$ is an unavoidable and mathematically rigorous consequence of Landauer's principle.

In an arbitrary learning system described by trainable variables q and nontrainable variables x, such that nontrainable variables undergo stochastic dynamics and trainable variables undergo learning dynamics. In the limit when the nontrainable variables x have already equilibrated, but the trainable variables q is still in the process of learning, the conditional probability distribution $p(x|q)$ over nontrainable variables x can be obtained from the maximum entropy principle whereby Shannon (or Boltzmann) entropy.

The principle of least action ensures a minimal energy conformation for objects moving in space, but intelligent systems optimize

their action repertoire between the past and the future (Deli et al., 2020). Intellect then is a temporal orientation—increasing future freedom of action. Therefore, intellect is a highly personal consideration of short and long-term consequences of actions. What amount of learning is required from the current knowledge p to reach knowledge q? When the distribution does not change with the change in the model parameters, the cost function is the Kullback-Leibler divergence (D_{KL}), a type of statistical distance and represented as a surprise between the current output and the expected output [12, 13].

In the endothermiccase, the discrete D_{KL} is defined as follows:
X = discrete random variable in brain signal space X.
p(x) ≥ 0, q(x) > 0 = probability distribution of x, where
p(x) = probability distribution of observed signal x, q(x) = probability distribution of estimated signal x

$$D_{KL}\big((p(x) \parallel q(x)\big) = \sum_{x \in X} p(x) \ln \frac{p(x)}{q(x)} \geq 0 \quad (1)$$

where ln $\frac{p(x)}{q(x)}$ is the novelty of the information. Learning depends on $D_{KL} \geq 0$ and when q = p, no learning
occurs.
Instead of asking the amount of learning taking place between p and q, we can look for the cognitive updating caused by an input, F (the free energy). S is the entropy, and T is the social temperature as defined by δQ

frequencies, dS = $\frac{\delta Q}{T}$, therefore

F = <F> - TS (2)

where <F> is the expected meaning (intellect). The system's free energy is proportional to the information's surprise value relative to expectation. As the system moves toward equilibrium, or high entropy, the system absorbs free energy, minimizing free energy.

The system evolves through a Markov process. F(q) – F(p) = kT D_{KL} (4)
where F(q) is the free energy for q and F(p) is the free energy in p.

Because the system's free energy is proportional to the information's surprise we can express the energy requirement of learning from Eq. (2),

$$\frac{dp_i}{dt} = (F_i - < F >)\, p_i \quad (5)$$

$$\frac{d}{dt} D_{KL} = -\sum_i (F_i - < F >)\, q_i \quad (6)$$

where $\sum_i (F^i - < F >) pi$ is the average "relevance" of an incoming stimulus, D_{KL} is how much information is left to learn by going from p to q. Equation 6 shows the relevance compared to the equilibrium position p_i. Information with greater surprise changes the brain more significantly. Optimal learning occurs if the "novelty" of information is relevant but manageable [113, 114].

(A-B) Attention Compresses Incoming Sensory Information (Adiabatic Compression)

The information input is an isothermal heat transfer via a preferred gradient. Because the average energy transferred per interaction is the connection strength times the frequency, the average connections' strength (information transfer density) between neurons correlates with temperature. The frequencies reflect the computational limit and the energy need of synaptic changes. Evoked activation channels compress information, reducing the input size and creating a high-density manifold [16]. The bottleneck of the next hidden layer constitutes as a compression of the information.
ΔG=ΔH−TΔS ΔH>0 TΔS>0

(B-C) Rejection of Inconsequential Information (Isothermal Compression)

Data compression rejects noise (waste heat), which collapses entropy. The synaptic map condenses information into a holographic, temporal representation. Memory formation is an orthogonal transformation satisfying Landauer's principle. The erasure cost is the work value of information [4, 5].
Filtering the information f before the subsequent hidden layer reduces the number of microstates.

(C-D) The spreading of the relevant signal (isothermal heat rejection)

Lower frequencies represent a response. The slowdown lowers the temperature, which corresponds to a temporal expansion.
The microstate loss decreases temperature and permits complexity increase. ΔH>0 TΔS>0

(D-A) Resting (isentropic) expansion

The brain's autonomous regulation restores the high entropy resting state, which prepares the system for new input. Thus, the resting condition is a thermodynamic requirement of the cyclic process. The information input increases entropy.

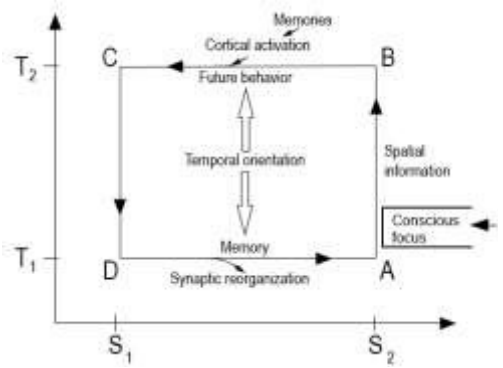

Figure 2: The reversed Carnot cycle representation of the brain evoked cycle

The cycle operates between the high frequency, evoked state (T2), and the resting state (T1); the horizontal axis represents entropy.

First, stimulus increases frequencies, compressing information as a function of conscious focus (AB). The electric flow reorganizes the synaptic potentials based on top-down memory (BC). Third, electric flow reversal formulates a response (CD). Finally, self-regulation recovers the high entropy resting state, readying the brain to receive new data (DA).

Discussion

Intelligent computation is the transformation of information into representation and meaning. The erasure cost (the work value of information) is the string's best compression [24]. Nevertheless, representation depends on a second relaxation step, manifested as an irreversible slowing down [29]. In cyclic processes, the relaxation time is on the order of the operation time.

Slowing down during gradient descent is a temporal expansion which reduces the error [30]. Therefore, orthogonal transformation stretches the spatial signal across the time domain. Orthogonal transform might arise via infinitesimal rotations by the hidden layers, which thus condense information into a holographic state of the network connectivity [31]. Percolation theory where the increasing number of links in a network generates globally connected node clusters, might serve as an analogy [32].

The view here is that this "erasure phase" is a "reset phase," which enriches mental complexity. The reset phase permits the system to acquire new information and begin the next cycle. The transformation is information storage is analogous to a temperature decrease via "phase transition," the phase energy requirement limits achievable computational efficiency. The system's organization represents a mental model which makes intelligent decision-making possible.

Landauer's principle dictates that information enrichment generates temperature in closed systems, but information processing reduces temperature and drives the cycle. Information erasure requires work on the system, which is dissipated as heat to the environment. The necessary amount of work is determined by our uncertainty about the system — the more we know about the system, the less it costs to 'erase' it. Although these processes' exact mechanism is not understood, information transformation can explain intelligent computation in animals and AI systems.

A Siamese network consists of two parallel and similar output vectors with the same weights [33]. In contrast, a hierarchic network consists of an input system and a relatively stable, orthogonal field [15]. The stability of the field representation can influence recognition and decision-making in a top- down manner. Therefore, the output represents the hidden layer outcome, influenced by the field vector in a top-down manner (Figure 3). Ultimately, the cycle's highly fluid operation engenders a gradual field evolution.

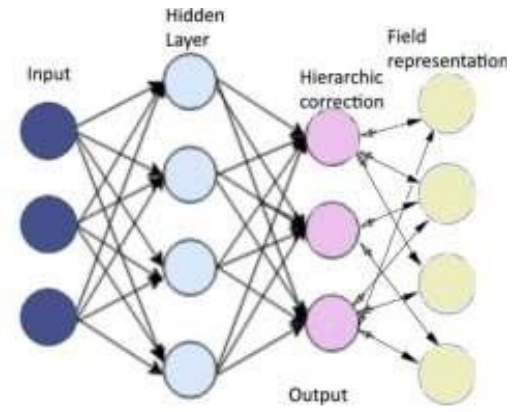

Figure 3: Hierarchic representation of the network the input arrives on the left. Information processing occurs in a highly fluid thermodynamic cycle. The field influences the representation via a top-down regulation. Output on the right represents the response.

Conclusion

We sought a deeper understanding of AI's "black box" by examining sensory interaction as an energy- information exchange. Information processing occurs in two hierarchical networks. The sensory system records the physical world's spatial organization by forming a closed thermodynamic cycle as a resting state centered temporally structured discrete processing. Therefore, the resting state readies the system to absorb new information.

The reversed Carnot cycle operating between two information conditions can model intelligent computation. First, information compression eliminates noise, and expands the system, ultimately restoring the initial conditions. The entropy maximizing self-regulation boosts future freedom of action forming an intellectual evolution. The first network's thermodynamic cycle processes information and updates the second network, the so-called field. The field accumulates memory to ensure intelligent decision-making.

Given enough time, systems with memory capacity accumulate intellect via endothermic cycles. Our thermodynamic considerations might provide a deeper understanding of AI's "black box."

Funding

This research received no external funding.

Conflict of interest

On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

1. Bubic, A., Von Cramon, D. Y., & Schubotz, R. I. (2010). Prediction, cognition and the brain. Frontiers in human neuroscience, 4, 25.
2. Fingelkurts, A. A., & Fingelkurts, A. A. (2015). Operational architectonics methodology for EEG analysis: theory and re-

sults. Neuromethods 91: 1-59.
3. Deli, E., Peters, J. F., & Tozzi, A. R. T. U. R. O. (2018). The thermodynamic analysis of neural computation. J Neurosci Clin Res 3, 1, 2.
4. Deli, E. (2021). Thermodynamic Computation In Self-Evolving Systems. Gerlingen: GmBh. doi, 10.
5. Deli, E., Peters, J., & Kisvárday, Z. (2021). The thermodynamics of cognition: a mathematical treatment. Computational and Structural Biotechnology Journal, 19, 784-793.
6. Maass, W., Papadimitriou, C. H., Vempala, S., & Legenstein, R. (2019). Brain computation: a computer science perspective. In Computing and Software Science (pp. 184-199). Springer, Cham.
7. Alain, G., & Bengio, Y. (2016). Understanding intermediate layers using linear classifier probes. arXiv preprint arXiv:1610.01644.
8. Wissner-Gross, A. D., & Freer, C. E. (2013). Causal entropic forces. Physical review letters, 110(16), 168702.
9. Cerezo, S. H., & Ballester, G. D. (2018). Fractal AI: A fragile theory of intelligence. arXiv preprint arXiv:1803.05049.
10. Cerezo, S. H., Ballester, G. D., & Baxevanakis, S. (2018). Solving Atari Games using fractals and entropy. arXiv preprint arXiv:1807.01081.
11. Deli, E. (2020). Can the fermionic mind hypothesis (FMH) explain consciousness? the physics of selfhood. Activitas Nervosa Superior, 62(2), 35-47.
12. Ryan, R. M., & Deci, E. L. (2017). Self-determination theory: Basic psychological needs in motivation, development, and wellness. Guilford Publications.
13. Tsao, A., Sugar, J., Lu, L., Wang, C., Knierim, J. J., Moser, M. B., & Moser, E. I. (2018). Integrating time from experience in the lateral entorhinal cortex. Nature, 561(7721), 57-62.
14. El-Kalliny, M. M., Wittig, J. H., Sheehan, T. C., Sreekumar, V., Inati, S. K., & Zaghloul, K. A. (2019). Changing temporal context in human temporal lobe promotes memory of distinct episodes. Nature communications, 10(1), 1-10.
15. Libby, A., & Buschman, T. J. (2021). Rotational dynamics reduce interference between sensory and memory representations. Nature neuroscience, 24(5), 715-726.
16. Saaty, T. L., & Vargas, L. G. (2017). Origin of neural firing and synthesis in making comparisons. European Journal of Pure and Applied Mathematics, 10(4), 602-613.
17. Buzsáki, G., Logothetis, N., & Singer, W. (2013). Scaling brain size, keeping timing: evolutionary preservation of brain rhythms. Neuron, 80(3), 751-764.
18. Déli, E., & Kisvárday, Z. (2020). The thermodynamic brain and the evolution of intellect: the role of mental energy. Cognitive neurodynamics, 14(6), 743-756
19. Deli, E., Peters, J., & Kisvárday, Z. (2021). The thermodynamics of cognition: a mathematical treatment. Computational and Structural Biotechnology Journal, 19, 784-793.
20. Cox, R. T. (1979). Of inference and inquiry, an essay in inductive logic. The maximum entropy formalism, 119-167.
21. Fry, R. L. (2017). Physical intelligence and thermodynamic computing. Entropy, 19(3), 107.
22. Gao, X., & Duan, L. M. (2017). Efficient representation of quantum many-body states with deep neural networks. Nature communications, 8(1), 1-6.
23. Shwartz-Ziv, R., & Tishby, N. (2017). Opening the black box of deep neural networks via information. arXiv preprint arXiv:1703.00810.
24. Baumeler, Ä., & Wolf, S. (2019). Free energy of a general computation. Physical Review E, 100(5), 052115.
25. Sakai, J. (2020). How synaptic pruning shapes neural wiring during development and, possibly, in disease. Proceedings of the National Academy of Sciences, 117(28), 16096-16099.
26. Tadros, T., Krishnan, G., Ramyaa, R., & Bazhenov, M. (2019). Biologically inspired sleep algorithm for increased generalization and adversarial robustness in deep neural networks. In International Conference on Learning Representations.
27. Landauer, R. (1961). Irreversibility and heat generation in the computing process. IBM journal of research and development, 5(3), 183-191.
28. Landauer R. [Journal] // Physics Today. - 1991. - Vol. 44. - p. 23.
29. Bennett, C. H. (1982). The thermodynamics of computation—a review. International Journal of Theoretical Physics, 21(12), 905-940.
30. Tozzi, A., & Peters, J. F. (2016). A topological approach unveils system invariances and broken symmetries in the brain. Journal of Neuroscience Research, 94(5), 351-365.
31. Lade, S. J., & Gross, T. (2012). Early warning signals for critical transitions: a generalized modeling approach. PLoS computational biology, 8(2), e1002360.
32. Broadbent, S. R., & Hammersley, J. M. (1957, July). Percolation processes: I. Crystals and mazes. In Mathematical proceedings of the Cambridge philosophical society (Vol. 53, No. 3, pp. 629-641). Cambridge University Press.
33. Chicco, D. (2021). Siamese neural networks: An overview. Artificial Neural Networks, 73-94.