Review | **Open Access**

ZENtropy

Liu *et al.* *ZENtropy* **2026**, *1*, 2 **DOI:**10.20517/zentropy.2025.02

From neurons to algorithms: perspectives on human intelligence and artificial intelligence

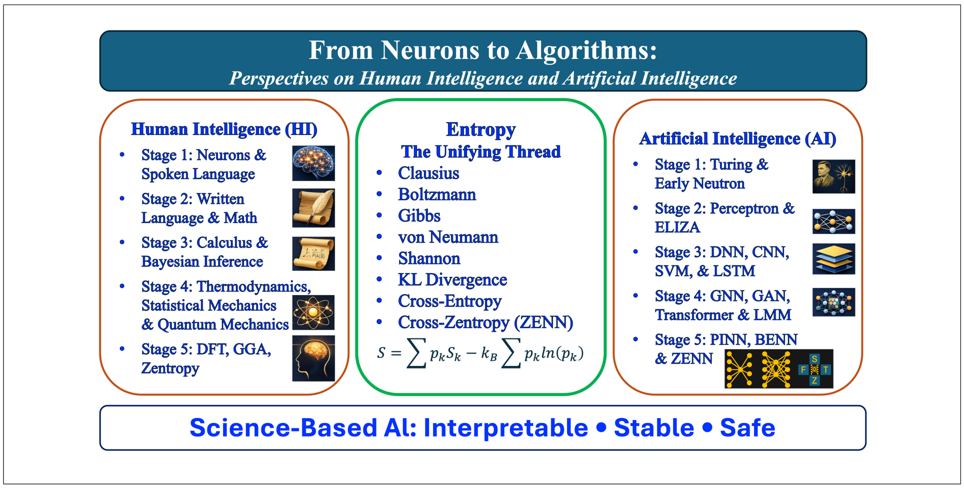
Zi-Kui Liu[1], Wenrui Hao[2], Bing Li[3]

Keywords:
Artificial intelligence, human intelligence, DNN, LLM, PINN, BENN, ZENN

Citation: Liu, Z. K.; Hao, W.; Li, B. From neurons to algorithms: perspectives on human intelligence and artificial intelligence. *ZENtropy* **2026**, *1*, 2. https://dx.doi.org/10.20517/zentropy.2025.02

Received: 31 Dec 2025
First Decision: 24 Apr 2026
Revised: 23 May 2026
Accepted: 8 Jul 2026
Published: 23 Jul 2026

Academic Editor: Weihua Xu
Copy Editor: Shu-Yuan Duan
Production Editor: Shu-Yuan Duan



Abstract

In this review article, we compare and contrast the evolution of human intelligence (HI) and artificial intelligence (AI). We show that both have progressed along strikingly parallel trajectories: from the development of basic neurons to simple and then complex neural networks; from the emergence of language to increasingly sophisticated mathematics; and from rudimentary physical principles to comprehensive physical laws. While modern AI systems demonstrate remarkable capabilities in pattern recognition, they still lack interpretability, physical grounding, and scientific rigor. To transition AI from a primarily data-driven tool into a science-based engine, we advocate embedding foundational principles of statistics and thermodynamics directly into AI frameworks. Towards this objective, we have recently developed the Belted and Ensembled Neural Networks (BENN) and the Zentropy-Enhanced Neural Networks (ZENN). BENN incorporates sufficient dimension reduction, enabling models to learn low-dimensional structures that preserve statistical information. ZENN integrates quantum mechanics and statistical mechanics into neural architectures, providing theory-driven models capable of reasoning across multiple scales while maintaining transparency and physical interpretability. We anticipate that such science-rooted AI frameworks could enhance not only the predictive reliability of next-generation AI models, but also their safety, by constraining learning within

[1]Department of Materials Science and Engineering, The Pennsylvania State University, Pennsylvania, PA 16802, USA.
[2]Department of Mathematics, The Pennsylvania State University, Pennsylvania, PA 16802, USA.
[3]Department of Statistics, The Pennsylvania State University, Pennsylvania, PA 16802, USA.

Correspondence to: Prof. Zi-Kui Liu, Department of Materials Science and Engineering, The Pennsylvania State University, Pennsylvania, PA 16802, USA. E-mail: prof.zikui.liu@psu.edu

established scientific principles, ensuring rigor, and improving robustness.

INTRODUCTION

Artificial intelligence (AI) has emerged as a transformative force across virtually every discipline, reshaping how we analyze data, model complex systems, and make decisions. Its capabilities in pattern recognition, prediction, and automation are well established, driving breakthroughs in fields ranging from healthcare and materials science to finance and climate modeling. However, despite these successes, most AI models remain fundamentally empirical, relying on massive datasets and opaque architectures with limited interpretability and weak connections to underlying scientific principles. This disconnect poses challenges for safety, reliability, transparency, and extrapolation beyond training regimes, which are critical requirements for scientific discovery and engineering design.

To address these limitations, the next generation of AI must go beyond data-driven heuristics and incorporate the rigor of mathematics, statistics, and physics into its core architectures. In the present review article, we discuss the evolution of human intelligence (HI) and contrast it with the evolution of AI to understand, anticipate and hopefully manage the trajectory of AI development. The emerging frameworks developed by the present authors are presented as examples of these efforts, including the Belted and Ensembled Neural Network (BENN)[1], which embeds sufficient dimension reduction into deep learning, and Zentropy-Enhanced Neural Networks (ZENN)[2], which integrate bottom-up quantum mechanics and top-down statistical mechanics. BENN and ZENN exemplify this shift toward theory-based AI by embedding physical laws, thermodynamic constraints, and statistical foundations into AI models, which have the potential to not only enhance their predictive power, but also transform them into interpretable, scientifically grounded engines capable of advancing discovery and predictions with safety bounded by fundamental physical laws.

This article's goal is to highlight a conceptual trajectory in the evolution of intelligence and to present BENN and ZENN as illustrative frameworks that exemplify how statistical and thermodynamic principles may be embedded into AI. The remainder of this paper is organized as follows. We first examine the evolution of human intelligence (**Section HUMAN INTELLIGENCE**) and draw parallels with artificial intelligence (**Section ARTIFICIAL INTELLIGENCE**). We then introduce entropy as a unifying concept linking these developments (**Section ENTROPY IN AI**), before presenting BENN and ZENN as representative frameworks (**Section NEW FRONTIERS: BENN AND ZENN**). Finally, we discuss implications for AI safety, stability, and future directions (**Section THERMODYNAMIC STABILITY, AI SAFETY, AND THE ROLE OF ZENTROPY**).

HUMAN INTELLIGENCE

Human intelligence (HI) has evolved through successive layers of abstraction and complexity, each layer adding structure and interpretability to the way information is processed as approximately shown in Table 1 and Figure 1, respectively. It is noted that the x-axis in Figure 1 is plotted in logarithmic scale. Figure 1 does not include the formation of neurons due to its long period of time and depicts approximately four stages with the coming fifth stage.

At the HI foundation lie neurons as biological units that form vast, interconnected networks enabling perception, learning, and cognition. These networks, comprising more than 100 trillion synaptic connections among roughly 86 billion neurons in the human brain[3-5], gave rise to spoken language, resulting in a transformative leap that revolutionized interpersonal communication and the exchange of ideas as shown as

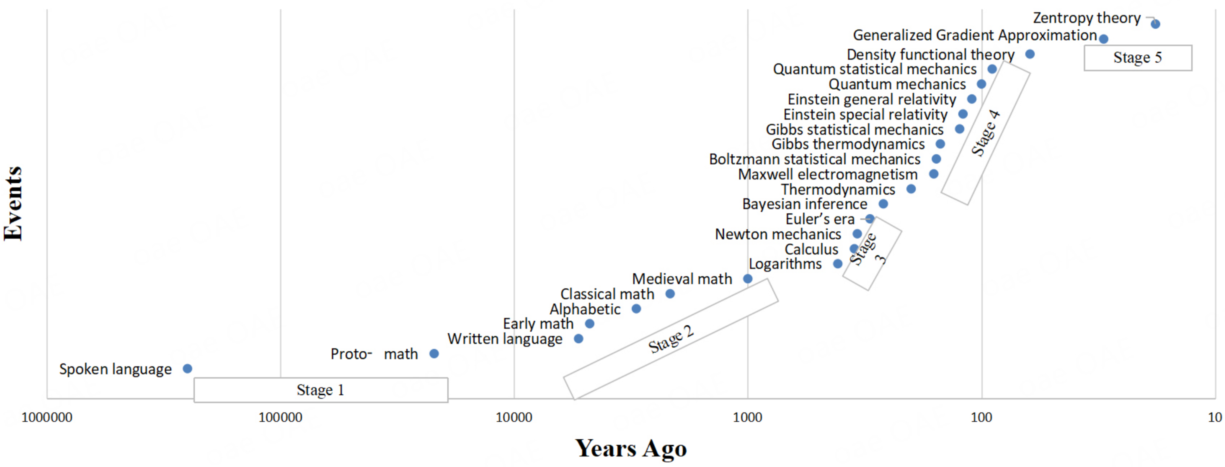


Figure 1. Evolution of human intelligence with the x-axis for years before 2026 plotted in logarithmic scale.

Table 1. Evolution of human intelligence

Human intelligence	Years ago	Years ago (figure)	Year
Neurons	~ 600,000,000	600,000,000	-599,997,974
Spoken language	200,000-300,000	250,000	-247,974
Proto-math	~ 22,000	22,000	-19,974
Written language	~ 5,400-5,200	5,300	-3,274
Early math	~ 5,000-4,500	4,750	-2,724
Alphabetic	~ 3,200-2,800	3,000	-974
Classical math	~ 2,300-2,000	2,150	-124
Medieval math	~ 1,200-800	1,000	1,026
Logarithms	~ 412	412	1,614
Calculus	~ 351-341	350	1,676
Newton mechanics	~ 339	340	1,687
Euler's era	~ 319-243	300	1,726
Bayesian inference	~ 263-214	263	1,763
Thermodynamics	~ 202-156	200	1,826
Maxwell electromagnetism	~ 165-153	160	1,865
Boltzmann statistical mechanics	~ 156-124	156	1,870
Gibbs thermodynamics	~ 202-156	150	1,876
Gibbs statistical mechanics	~ 124	124	1,902
Einstein special relativity	~ 121	120	1,905
Einstein general relativity	~ 111	110	1,915
Quantum mechanics	~ 126-99	100	1,926
Quantum statistical mechanics	~ 102-85	90	1,936
Density functional theory	~ 62-61	62	1,964
Generalized Gradient Approximation	~ 40-38	30	1,996
Zentropy theory	~ 18-4	18	2,008

Stage 1 of HI in Figure 1. The emergence of language allowed humans to share experiences, coordinate actions, and transmit cultural knowledge, laying the groundwork for social organization and collective progress, including simple math such as tally marks.

The advent of written language marked another milestone, Stage 2 of HI as shown in Figure 1, enabling knowledge to be preserved, accumulated, and transmitted across generations, thus amplifying the reach and durability of human thought. This intellectual progression ultimately culminated in the development of scientific reasoning, expressed through more complex mathematical formulations.

Stage 3 started from the development of logarithms about 400 years ago, followed by significant development in mathematics, particularly calculus, Newton mechanics, and Bayesian inference. The first industrial revolution marked the beginning of Stage 4 of HI from macroscopic observations to microscopic understanding including the establishment of thermodynamics, statistical mechanics, special and general relativity, and quantum mechanics.

Figure 1 further indicates that density functional theory (DFT)[6,7] is leading a new Stage 5 of HI. DFT defines the ground-state configuration of a system at zero Kelvin with its energy at minimum determined by a unique electronic structure and electron density distribution and provides a practical solution of quantum mechanics. The DFT computation accuracy was dramatically improved by generalized gradient approximation (GGA)[8]. Prior to DFT and GGA, human knowledge was primarily accumulated through observations, measurements, interpretations, and empirical correlations. DFT + GGA enabled complete prediction from quantum mechanics under several well-defined approximations without additional external inputs. By integrating bottom-up predictions from DFT + GGA for both ground-state and excited-state configurations with top-down Gibbs statistical mechanics, zentropy theory[9] demonstrated remarkable agreement with experimentally observed emergent behaviors in magnetic materials[10,11], including phase transition, singularity, and negative thermal expansion[12], without empirical models and fitting parameters, providing a practical solution for quantum statistical mechanics[13]. This stage marks a new era in which new knowledge can be created quantitatively and accurately by quantum-based scientific computations and then validated by measurements for practical applications. Consequently, the rate of human knowledge accumulation will be exponentially faster and more accurate than ever before.

ARTIFICIAL INTELLIGENCE

The HI historical trajectory illustrates a fundamental principle: intelligence advances by embedding theory into increasingly sophisticated representations, driving humanity from raw perception to predictive understanding. Modern AI architectures reflect this evolutionary trajectory of intelligence, echoing the layered complexity of the human intelligence as approximately shown in Table 2 and Figure 2, respectively.

Similar to HI, the development of AI can be broadly grouped into four stages with the potential fifth stage as marked in Figure 2. Stage 1 and Stage 2 can be attributed to the initiation of computation and neural networks that ended in the so called AI winter[14]. Stage 3 represents the development of Deep Neural Networks (DNNs)[15], Convolutional Neural Networks (CNNs)[16,17], and Long Short-Term Memory (LSTM)[18], while Stage 4 introduces the Graph Neural Networks (GNNs)[19], Large Language Models (LLMs)[20,21], and Physics-Informed Neural Networks (PINNs)[22].

DNNs[15] attempt to emulate aspects of human perceptual intelligence by constructing hierarchical systems of artificial neurons that extract increasingly abstract features from raw data. As specialized variants of DNNs, CNNs[16,17], LSTMs[18], and GNNs[19] each capture different structural aspects of the world from spatial locality in images, temporal continuity in sequences, to relational connectivity in graphs, which roughly parallel how human sensory systems process visual scenes, temporal patterns, and relational structures. Like early HI, these networks learn almost entirely from statistical correlations and require large quantities of labeled data. Progress in attention mechanisms, self-supervised learning, and hybrid architectures helps DNNs move

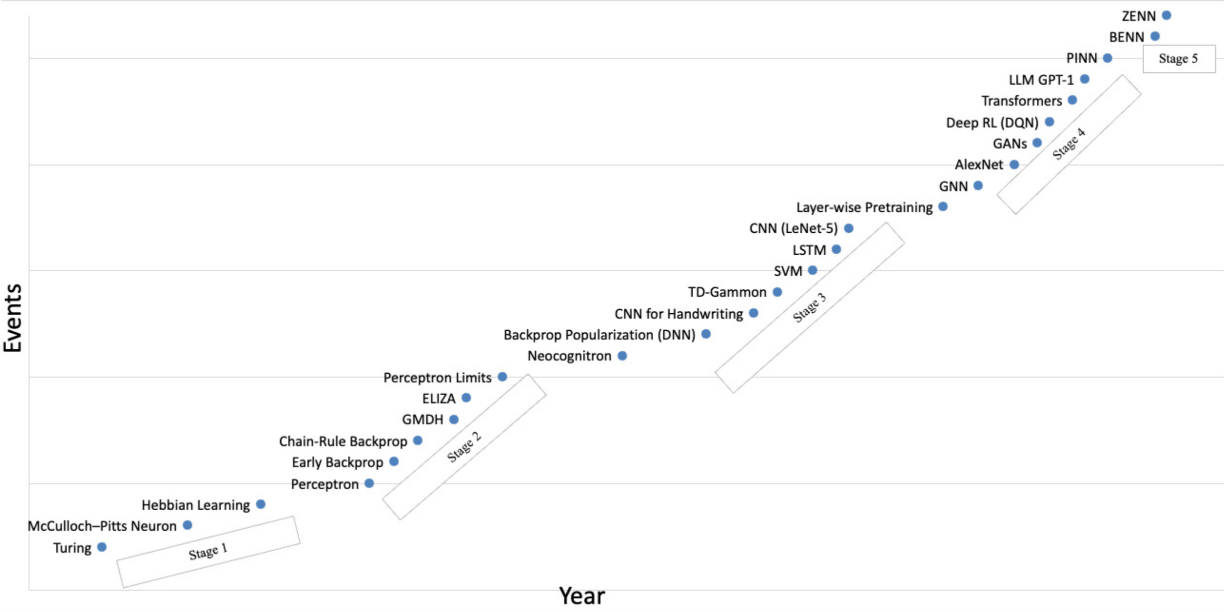


Figure 2. Evolution of artificial intelligence with the x-axis for year plotted in logarithmic scale. DNN: Deep Neural Network; CNN: Convolutional Neural Network; TD-Gammon: SVM: Support-Vector Network; LSTM: Long Short-Term Memory; GNN: Graph Neural Network; GAN: Generative Adversarial Network; DQN: LLM: Large Language Model; PINN: Physics-Informed Neural Network; BENN: Belted and Ensembled Neural Network; ZENN: Zentropy-Enhanced Neural Network.

closer to the flexibility and efficiency seen in HI by enabling longer-range context integration, more robust representation learning, and more human-like abstraction.

LLMs[20,21] further advance toward human linguistic and conceptual intelligence by operating on tokens that represent words or sub-word units, embedding them into high-dimensional representation spaces, and using transformers to model semantic and contextual relationships across extensive text sequences. Their ability to generate coherent language, analyze complex concepts, and generalize across domains parallels HI's facility with abstraction, analogy, and reasoning through symbolic expressions in spoken and written languages. At the same time, LLMs still lack the grounded understanding that characterizes HI: they do not possess embedded physical models, causal reasoning engines, or the experiential grounding that humans use to connect symbols to the physical world. Emerging techniques such as retrieval augmentation[23], reasoning verifiers[24-26], and tool-use frameworks[27] push LLMs toward forms of externalized cognition reminiscent of how humans consult references, perform calculations, or check logical consistency, thereby narrowing (though not eliminating) the gap between statistical language modeling and more structured reasoning. LLMs have enabled machines to process and generate human language with unprecedented fluency and contextual understanding and transformed natural language processing, powering applications from conversational agents to automated reasoning systems[20,21].

As the frontier advances, AI has begun to integrate the rigor of physical sciences. PINNs[22,28,29] and Deep Ritz method[30] bring AI closer to human scientific intelligence by integrating symbolic reasoning, mathematical structure, and physical understanding. By embedding partial differential equations (PDEs), conservation laws, and boundary conditions directly into their training objectives, PINNs[22,28,29] mimic the way HI imposes physical constraints on reasoning, ensuring that predictions remain consistent with established laws rather than purely data-driven correlations. While these methods struggle with optimization stiffness, multiscale structures, and complex geometries, ongoing developments, such as domain decomposition, adaptive weighting, and hard constraints, improve their reliability and scalability as a critical step toward

Table 2. Evolution of artificial intelligence

Year	Short Name	Milestone
1936	Turing	On Computable Numbers; foundations of computability theory
1943	McCulloch-Pitts Neuron	First artificial neuron model by McCulloch & Pitts
1949	Hebbian Learning	Hebb introduces synaptic learning rule
1958	Perceptron	Rosenblatt's probabilistic model for information storage and organization
1960	Early Backprop	Kelley publishes early continuous backpropagation formulation
1962	Chain-Rule Backprop	Dreyfus refines backprop using the chain rule
1965	GMDH	Ivakhnenko & Lapa develop early multi-layer (deep) training (GMDH)
1966	ELIZA	First chatbot; rule-based natural language system
1969	Perceptron Limits	Minsky & Papert show limits of single-layer nets; AI winter
1979	Neocognitron	Fukushima introduces CNN precursor with hierarchical vision layers
1986	Backprop Popularization (DNN)	Rumelhart, Hinton & Williams popularize backprop
1990	CNN for Handwriting	LeCun applies CNNs to handwriting/zip code recognition
1992	TD-Gammon	Tesauro's RL system achieves championship-level backgammon
1995	SVM	Support-Vector Networks
1997	LSTM	Hochreiter & Schmidhuber introduce Long Short-Term Memory
1998	CNN (LeNet-5)	Gradient-based learning applied to document recognition (LeNet-5)
2006	Layer-wise Pretraining	Hinton, Osindero & Teh introduce greedy pretraining for deep nets
2009	GNN	Graph Neural Network Model
2012	AlexNet	ImageNet breakthrough with deep CNNs.
2014	GANs	Goodfellow introduces Generative Adversarial Networks.
2015	Deep RL (DQN)	Human-level control through deep reinforcement learning.
2017	Transformers	"Attention Is All You Need" introduces Transformer.
2018	LLM GPT-1	First modern LLM; introduced large-scale Transformer pretraining.
2020	PINN	Physics-Informed Neural Networks framework.
2024	BENN	Belted and Ensembled Neural Network.
2025	ZENN	Zentropy-Enhanced Neural Network; thermodynamics-inspired.

trustworthy scientific AI. Neural operators[31-34] extend this trajectory by learning mappings between entire function spaces, enabling rapid simulation across parameter ranges in a way that resembles how HI generalizes physical intuition from a small set of principles to many scenarios. Together, these models represent an evolution from pattern perception (DNNs) to linguistic abstraction (LLMs) to physically grounded reasoning (PINNs), collectively forming a pathway toward AI systems that reflect increasingly comprehensive aspects of HI.

ENTROPY IN AI

Having outlined the parallel evolution of HI and AI, we now turn to entropy as the central concept that connects these trajectories. From the HI and AI evolutions shown in Figures 1 and 2 and discussions in above sections, it seems that Stage 4 in HI is conspicuously missing in AI. Evidently, this could not be true if the evolution trajectory of AI follows that of HI as AI is developed by humans after all. The connection is in terms of one of the most profound and unifying concepts in science: entropy, devised in thermodynamics to formalize irreversibility and limits by Clausius[35], cast by statistical mechanics as counting microstates under constraints by Boltzmann[36] and Gibbs[37], and transformed by information theory into a quantitative measure of uncertainty that bridges physics and communication by Shannon[38].

Table 3. Historical through-line of entropy in AI

Year	Explicit entropy milestone in AI
1948	Shannon entropy as average surprisal; foundation of information theory
1951	Kullback-Leibler divergence formalized as information for discrimination
1957	Jaynes maximum-entropy inference linking statistical mechanics to information theory
1986	Entropy as information gain in structure learning
1990	Cross-entropy justified for neural classification
2008	Variational free energy and Kullback-Leibler divergence in modern inference
2014	KL-regularized generative modeling (VAE)
2018	Maximum-entropy deep reinforcement learning
2020	Entropy-controlled decoding in text generation
2022	KL-penalized alignment for LLMs (RLHF)
2025	Entropy as first-principles modeling in scientific AI (ZENN)

KL: Kullback-Leibler; AI: artificial intelligence; ZENN: Zentropy-Enhanced Neural Network.

In AI, this same quantity migrates from an abstract measure to a working instrument: it becomes the loss we minimize, the uncertainty we expose and calibrate, the constraint we enforce through governing equations, and the objective we optimize when trading accuracy against complexity and cost. The historical through-line of entropy is clearest if we read AI chronologically through its explicit uses of entropy as shown in Table 3.

Entropy in AI started with Shannon's foundation for measuring uncertainty[38] by introducing entropy as a measure of uncertainty in a single probability distribution. Kullback and Leibler extended this framework by defining a measure of relative information between two probability distributions, now called the Kullback-Leibler (KL) divergence[39]. This quantity expresses how much information is lost when one distribution is used to approximate another and thus generalizes Shannon's entropy from individual distributions to comparisons between distributions. Jaynes's maximum-entropy inference method[40] depends directly on both Shannon entropy and KL divergence. His principle chooses the distribution that maximizes Shannon entropy subject to constraints, and the formal derivation uses KL divergence to show that the maximum-entropy distribution is the one least distinguishable (in the KL sense) from a uniform or prior distribution under the given constraints. Jaynes cited the KL result as the basis for interpreting maximum-entropy inference as choosing the least biased, least informative distribution, bridging from physics to statistical inference[39,40].

Along AI trajectory, entropy has become both the criterion for choosing model structure and the objective for training probabilistic predictors. The backpropagation era[15,16] did not merely introduce depth; it elevated cross-entropy to a first-class training signal for classification and sequence prediction, made explicit by works that interpret neural outputs as normalized probabilities and justify cross-entropy as the proper objective for Bayesian *a-posteriori* estimation[41,42]. At roughly the same time, decision-tree learning formalized structure discovery as an entropy-reduction process-information gain, placing entropy at the heart of discrete model selection[43]. As probabilistic modeling matured, the variational methods cast learning and inference as the optimization of free-energy-like objectives in which KL divergence[39], the entropy-based gap between approximate and target distributions, regulates complexity while maximizing data fit[44]. Generative modeling made this explicit in the variational autoencoder, whose evidence lower bound trades reconstruction accuracy against a KL term that sculpts the latent space[45]. In reinforcement learning, the maximum-entropy formulation promotes policies that are simultaneously competent and suitably stochastic, preventing premature collapse and enabling robust exploration; the Soft Actor-Critic algorithm is a canonical deep instantiation of this principle[46].

The transformer era scales these ideas dramatically. LLMs are trained by minimizing cross-entropy at token level across vast corpora, and their inference procedures explicitly modulate sequence entropy to balance diversity with coherence through temperature scaling, top-k pruning, and nucleus sampling[47]. Alignment pipelines add another entropy-based ingredient by penalizing deviations from a reference distribution with a KL term during reinforcement learning from human feedback[25]. In comparison with Stage 4 in HI, this is precisely the thermodynamic turn: AI systems now treat entropy not only as a training loss but as a conserved budget and a controllable resource, exposing it when they must ask for information, constraining it when they must obey laws, and spending it when they must explore or generate.

In scientific AI, this trajectory opens the door to architectures and objectives that operationalize entropy the way thermodynamics intended: as the quantitative thread linking uncertainty, constraints, and work. The zentropy-guided view makes this explicit by connecting statistical sufficiency and free-energy landscapes to the practical choices an AI system makes about what to sense, retrieve, simulate, and enforce. The result is an entropy-aware loop that is faithful to the physics of information and to the scientific use of thermodynamic reasoning, exactly what Stage 4 anticipates.

Another important event in HI is Bayesian inference[48] at the end of Stage 3 as shown in Figure 1. It originally emerged as a framework for updating beliefs in the presence of uncertainty, long before entropy entered statistics. Classical Bayesian reasoning, as developed from Bayes and Laplace onward, formalized probability as a tool for learning from data but did not yet involve entropy explicitly. This connection was forged through the works by Shannon[38], Kullback and Leibler[39], and Jaynes[40] that enabled a deeper reinterpretation of Bayesian inference, thereby unifying Bayesian probability with thermodynamics and statistical mechanics. In this way, Bayesian inference sits conceptually between classical probability and modern entropy-based reasoning: it begins without entropy but later becomes one of the principal domains in which entropy, relative entropy, and maximum-entropy principles acquire their scientific meaning.

NEW FRONTIERS: BENN AND ZENN

Building on the role of entropy, we now introduce two frameworks, BENN and ZENN, the two directions that the present authors marched into, as concrete examples of embedding scientific principles into AI. In the first direction, BENN[1] is developed to expand the scientific foundations of AI by embedding statistical sufficiency directly into neural architectures. Rather than letting deep networks form high-dimensional latent representations without theoretical guarantees, the BENN framework[1] introduces principled sufficient dimension reduction (SDR) as an internal structural constraint. By doing so, BENN connects modern deep learning to one of the most fundamental constructs in statistics: identifying a low-dimensional subspace that captures all information in the predictor relevant to the response. BENN architecture is schematically shown in Figure 3[1].

At the core of BENN is its distinctive narrow “belt” layer, a structural bottleneck designed not merely for compression, as in autoencoders, but specifically for estimating the central subspace in which the conditional distribution of the response is preserved. This belt serves as a statistically grounded representational scaffold, enforcing a low-dimensional manifold that retains sufficiency even for complex nonlinear relationships. Complementing this structural design, BENN[1] employs an ensemble of response transformations, enabling the model to recover both conditional means and full conditional distributions, thereby supporting rigorous inference while maintaining the expressive flexibility of deep neural networks.

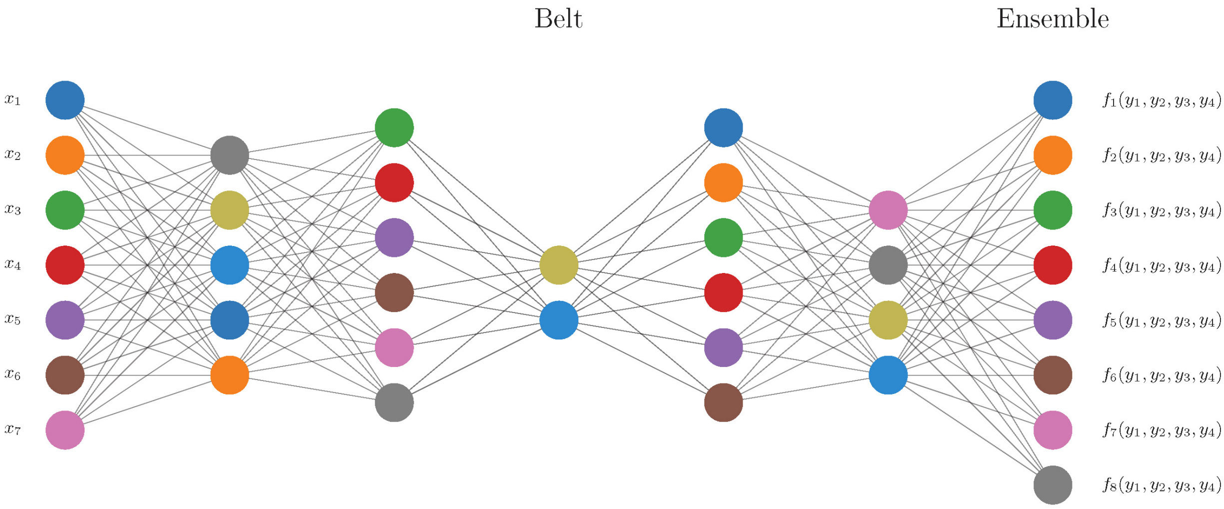


Figure 3. Structure of BENN for sufficient dimension reduction. The belt is the sufficient predictor, and the output is the ensemble containing the transformations of the responses. Redrawn by Yin Tang based on Ref.[1]. BENN: Belted and Ensembled Neural Network.

The output extracted from the belt plays a critical conceptual role: it summarizes the information scattered across thousands or millions of neural parameters into a compact set of sufficient statistics. This process parallels a hallmark of human cognition, i.e., our ability to distill the overwhelming complexity of the external world into concise abstractions, theories, and conceptual representations. In BENN[1], this abstraction is not heuristic but mathematically principled: the belt layer concentrates precisely those aspects of the input that govern the response, leaving behind extraneous variation. In this sense, BENN offers a neural analog to the way that scientific reasoning compresses vast empirical phenomena into low-dimensional, intelligible structures.

Together with its ensemble-based stabilization and its theoretical grounding in sufficient dimension reduction, BENN[1] exemplifies how statistics can be formally embedded into deep learning, transforming neural networks from purely correlational pattern-extractors into models capable of structured, evidence-preserving inference.

In the second direction, zentropy theory is embedded into the AI architecture. The central idea of zentropy theory[9-11] is that the total entropy of a system comprises two components: the Gibbs-Shannon entropy arising from the probability distribution across configurations and the intrinsic entropy residing within each configuration. Owing to the recursive nature of entropy, this intrinsic entropy ranges from zero for configurations without unconstrained degrees of freedom, i.e., quantum pure-state configurations, to the entire entropy of the system when the configuration is the system itself[49]. For realistic complex systems, the number of quantum pure-state configurations is astronomically large, making exhaustive enumeration impossible. Consequently, intrinsic entropy must be accounted for when evaluating both the total entropy of the system and the statistical weight (probability) of individual configurations, which are no longer pure-state configurations.

This is schematically shown in Figure 4. Figure 4A depicts the scenario with quantum pure-state configurations, thus without intrinsic entropy within each configuration, while Figure 4B includes an intermediate layer of configurations with each of them composed of more than one quantum pure-state configurations, resulting in intrinsic entropy in some intermediate configurations, such as the configurations defined by DFT. The total entropy of the system can be written as follows[49]

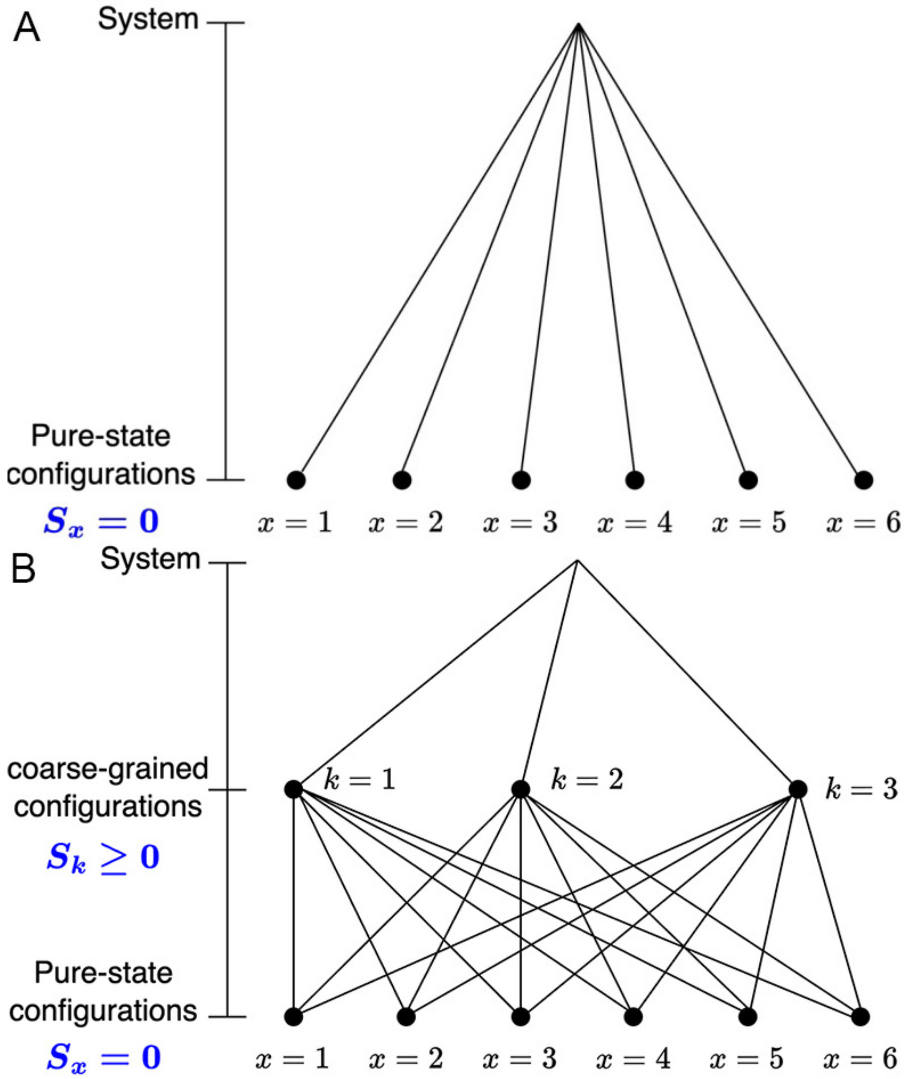


Figure 4. Schematic of recursive property of entropy in terms of probability tree diagrams with (A) p_x for the ungrouped quantum pure-state configurations (x) without intrinsic entroapy; and (B) $p_k = \sum_x p_{x|k}$ for the grouped intermediate configurations (k) with intrinsic entropy being $S_k = -k_B \sum_x p_{x|k} \ln\left(p_{x|k}\right)$ where $p_{x|k}$ is the conditional probability representing the contribution from configuration x to configuration k.

$$S = -k_B \sum_{x=1}^{6} p_x \ln\left(p_x\right) = -k_B \sum_{k=1}^{3} p_k \sum_x p_{x|k} \ln\left(p_{x|k}\right) - k_B \sum_{k=1}^{3} p_k \ln\left(p_k\right) = \sum_{k=1}^{3} p_k S_k - k_B \sum_{k=1}^{3} p_k \ln\left(p_k\right) \quad (1)$$

where k_B is the Boltzmann constant, and other quantities are defined in Figure 4 and its caption.

Zentropy theory, represented by Eq. 1, was first developed for magnetic crystals[10-12], where the intrinsic entropy of a configuration, i.e., its quantum entropy, is determined using DFT-based quantum mechanics with Fermi-Dirac statistics for electronic excitations and Bose-Einstein statistics for phonon contributions[50]. The theory was formalized in terms of entropy and statistics[51], though the term came later[9]. This framework establishes that entropy is not solely a property of the ensemble distribution but also an inherent property of each configuration itself, shaped by its internal quantum degrees of freedom. As a result, it becomes evident that the widely used cross-entropy in AI, which assumes configurations have no internal entropy, is fundamentally incomplete. A scientifically grounded AI must therefore revise or generalize cross-entropy to cross-zentropy in order to incorporate intrinsic entropy, ensuring that learning objectives reflect the full physical information content of states rather than only their distributional uncertainty.

Therefore, in this second direction, zentropy theory introduces two major conceptual advances grounded in top-down statistical mechanics and bottom-up quantum mechanics. The first advance is the replacement of conventional cross-entropy with cross-zentropy, allowing the learning objective to account for the total entropy of the system, not merely the Gibbs-Shannon component that describes the distribution among configurations, but also the intrinsic entropy residing within each configuration, as discussed above.

The second advance concerns the structural units that constitute a macroscopic system within an AI framework. In developing statistical mechanics, Gibbs[52] "imagined a great number of systems of the same nature, but differing in the configurations and velocities which they have at a given instant, and differing not merely infinitesimally, but so as to embrace every conceivable combination of configuration and velocities." This viewpoint establishes that a macroscopic system is fundamentally a statistical mixture of microscopic configurations. By analogy, AI systems should incorporate not only low-level primitives, such as neurons in DNNs[15-18] or tokens in LLMs[20,21], but also higher-level building blocks in the form of configurations, which capture the collective states underlying macroscopic observables.

Just as DFT provides a practical coarse-grained representation of quantum mechanics by integrating out the detailed behavior of electrons and phonons within each configuration, DNNs and LLMs can be viewed as coarse-grained representations of neurons and words. The macroscopic observables of a system then arise from the statistical mixture of such configurations, governed by statistical mechanicals principles. In thermodynamics, these observables correspond to derivatives of the free energy under prescribed external conditions[53-55].

In principle, one could imagine scaling DNNs or LLMs to astronomical sizes in an attempt to capture all quantum pure-state configurations directly. In practice, this is impossible, for the same reason that explicitly enumerating all quantum states in many-body systems is impossible: their number grows combinatorially without limit. Motivated by the success of DFT in quantum mechanics and zentropy theory in quantum statistical mechanics, we developed the ZENN AI framework[2], schematically shown in Figure 5, which incorporates both of these conceptual advances, i.e., cross-zentropy and configuration-level architecture, into a unified and physically grounded AI system. It is noted that the first improvement introduced by zentropy theory, the replacement of conventional cross-entropy with cross-zentropy, is primarily mathematical in nature and therefore not directly visible in Figure 5. This development, which generalizes the learning objective to include both configurational and intrinsic entropy, is discussed in detail in ZENN[2]. The second improvement is structural, redefining the building blocks of the AI framework to include configuration-level representations, and this architectural innovation is visually depicted in Figure 5.

To compute macroscopic observables, the Helmholtz energy of the system is expressed, following thermodynamic principles[53-55], as

$$F = U - TS = \sum_{k=1}^{W} p_k U_k - T\left[\sum_{k=1}^{W} p_k S_k - k_B \sum_{k=1}^{W} p_k \ln(p_k)\right] = \sum_{k=1}^{W} p_k F_k + k_B T \sum_{k=1}^{W} p_k \ln(p_k) \qquad (2)$$

where U, U_k, F and $F_k = U_k - TS_k$ are the internal energy and Helmholtz energy of the system and configuration k, respectively, T is temperature, and W is the number of configurations. Casting Equation 2 in the partition function formalism in Gibbs statistical mechanics, one obtains the following equations

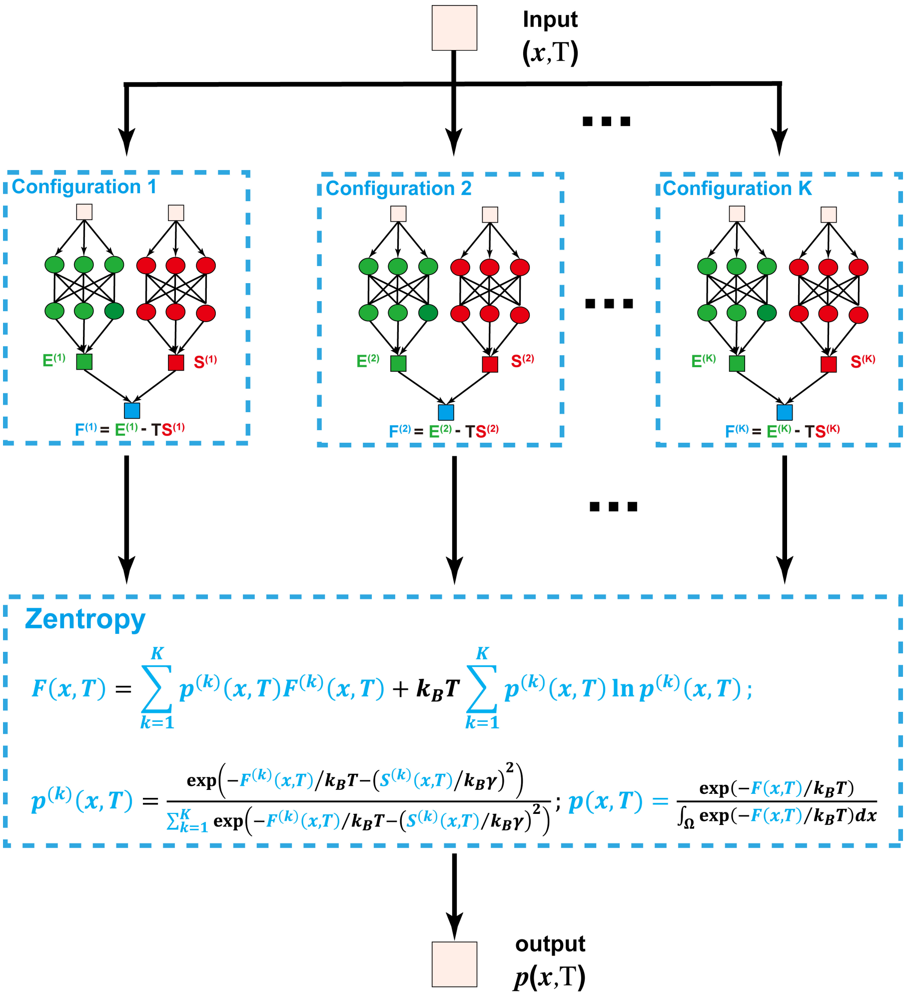


Figure 5. Schematic of ZENN AI framework. Reprint with Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND)[2]. ZENN: Zentropy-Enhanced Neural Network; AI: artificial intelligence.

$$Z = e^{-\frac{F}{k_B T}} = \sum_{k=1}^{W} Z_k = \sum_{k=1}^{W} e^{-\frac{F_k}{k_B T}} \tag{3}$$

and the probability of configuration k is:

$$p_k = \frac{Z_k}{Z} = e^{-\frac{F_k - F}{k_B T}} \tag{4}$$

where Z and Z_k are the partition functions of the system and configuration k, respectively, which was incorporated into the term “zentropy”, derived from the German word Zustandssumme, meaning “sum over states”[9,10].

In principle, any combination of configurations may be considered, provided they are “of the same nature”[52], meaning they share the same external constraints as the system, because of the recursive property of entropy illustrated in Figure 4. This recursive structure is not merely a mathematical curiosity; it enables a unified description of systems across scales, from quantum-level configurations to macroscopic thermodynamic ensembles, making cross-scale modeling both natural and rigorous. By uniting quantum-level precision with statistical-mechanics rigor and leveraging this recursive nature, zentropy theory and ZENN establish a genuinely cross-scale modeling framework. This framework links electronic structure,

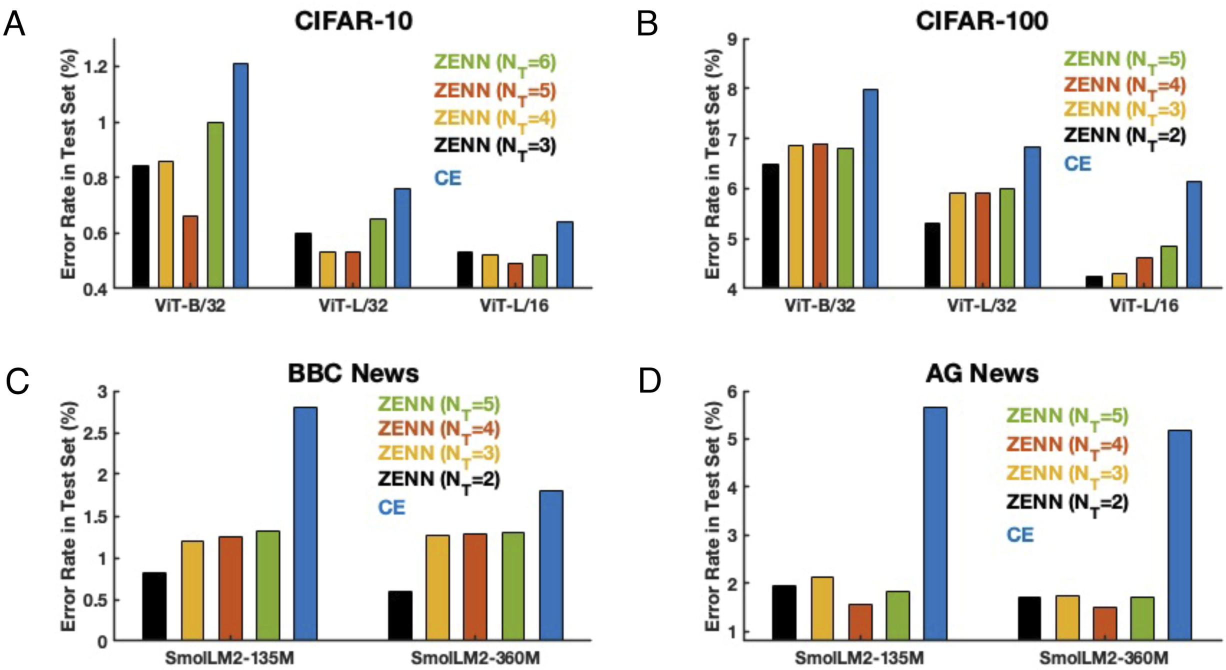


Figure 6. Error rate comparison between cross-entropy (CE) and cross-zentropy (ZENN) on both image and text datasets: (A) and (B) three ViT on CIFAR-10 and CIFAR-100 image datasets; (C) and (D) two language models (SmolLM) on BBC News and AG News text datasets. NT denotes the number of temperature modes. Reprint with Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND) [2]. ViT: Vision Transformer; ZENN: Zentropy-Enhanced Neural Network.

thermodynamic ensembles, and emergent macroscopic behavior, marking a transition toward science-based intelligence capable of reasoning across scales and predicting complex phenomena with unprecedented fidelity.

As shown schematically in Figure 5, ZENN implements this framework by representing the internal energy and intrinsic entropy of each configuration using separate DNNs tailored to the systems under study. The final model output is then computed using zentropy theory, ensuring that both configurational distributions and quantum-mechanical contributions are properly captured. ZENN models developed for image and text classification, using CIFAR-10 and CIFAR-100 for images and BBC News and AG News for text, consistently outperform state-of-the-art architectures, including Vision Transformers (ViT) for images and SmolLM language models for text, in both accuracy and robustness[2,56]. As summarized in Figure 6, with an optimally learned temperature, ZENN reduces relative error by approximately 20%-50% on CIFAR-10 and CIFAR-100 compared to ViT models, and by 60%-70% on BBC News and AG News compared to SmolLM. Moreover, ZENN built on small ViT or SmolLM backbones achieves error rates comparable to, or better than, those of their substantially larger counterparts.

The superiority of ZENN arises from its ability to model systems as ensembles of configurations, learning the intrinsic properties of each configuration rather than relying solely on superficial external features[2,9,57]. This ensemble-based perspective enables ZENN to generalize more effectively, handle uncertainty with greater fidelity, and yield interpretable insights into the underlying structure of data. In image classification, ZENN does not merely memorize pixel-level patterns; it captures the statistical relationships that govern object representation. In text classification, it identifies semantic structures and probabilistic dependencies that go beyond keyword matching. Unlike conventional architectures that primarily differentiate raw data through high-dimensional feature manipulation, ZENN explicitly models intrinsic entropy, enabling reliable extrapolation beyond the training domain and delivering smooth, physically meaningful derivatives crucial for sensitivity analysis and interpretability. These capabilities are particularly important for scientific and engineering applications, where model predictions must remain stable under varying conditions and where derivatives correspond to physical quantities such as forces, susceptibilities, or response functions.

By embedding thermodynamic principles into machine learning, ZENN transforms AI from a predominantly black-box predictor into a transparent, theory-guided engine for discovery. This integration bridges the long-standing gap between data-driven learning and science-based reasoning, paving the way for AI systems that are not only accurate and robust but also interpretable, reliable, and capable of supporting scientific innovation. In doing so, ZENN represents a step toward science-based paradigm for AI, in which physical and statistical principles are more explicitly embedded into model design to address complex problems across domains.

Furthermore, it is noted that independent studies across computer science and artificial intelligence[58-64] have consistently shown that zentropy-based feature selection outperforms leading state-of-the-art approaches, underscoring its promise as a broadly applicable and robust analytical framework. This conclusion draws strength from experiments conducted on more than eighty benchmark datasets that span a wide array of domains and data characteristics. These datasets include medically oriented collections such as Arrhythmia, Breast Cancer Wisconsin Diagnostic, Colon, and Leukemia, which pose challenges related to high dimensionality and limited sample sizes. They also encompass dynamic sensing and time-series tasks like Cardiotocography and Sonar, complex simulation data exemplified by Climate Model Simulation Crashes, and socio-economic decision-support datasets such as Nursery and South German Credit. Moreover, high-dimensional pattern-recognition and text-related datasets like Spambase, WarpAR10P, and Wine Quality White further demonstrate the breadth of contexts in which zentropy-based approaches have been evaluated.

What emerges from this extensive body of evidence[58-64] is a striking level of performance stability. Zentropy does not falter when confronted with heterogeneous feature types, noisy measurements, or substantial class imbalance. Instead, its core principle, capturing uncertainty across multiple interacting granular levels, allows it to extract meaningful structure when traditional feature selection methods struggle. Because it integrates coarse and fine information rather than compressing uncertainty into a single level, zentropy preserves subtle but informative distinctions that are often lost in conventional entropy-based or distance-based criteria. This multi-granular perspective is key to its success across such varied domains.

The consistency of these results further suggests that zentropy is not merely a high-performing technique tailored to specific problem families but rather a general mechanism that aligns naturally with the way complex datasets encode information. This is due to the universality of thermodynamics as Einstein wrote "A theory is the more impressive the greater the simplicity of its premises is, the more different kinds of things it relates, and the more extended is its area of applicability. Therefore the deep impression which classical thermodynamics made upon me. It is the only physical theory of universal content which I am convinced that, within the framework of the applicability of its basic concepts, it will never be overthrown"[65,66]. Its ability to maintain accuracy and interpretability across biomedical data, simulation science, socio-economic decision contexts, and high-dimensional recognition tasks points to a unifying strength: zentropy is well-suited for modern data landscapes where structural complexity, heterogeneity, and imbalance are the norm. Through its robust uncertainty modeling, it offers a reliable pathway for uncovering relevant features across diverse and challenging environments, reinforcing the view that zentropy's potential extends well beyond individual applications and holds genuine interdisciplinary value.

We note that the present results are still limited to selected benchmark datasets and model architectures. Further validation across broader domains is necessary and underway. In addition, alternative approaches, such as Bayesian deep learning, probabilistic programming, and other physics-informed or hybrid models, also aim to address interpretability and robustness. ZENN should therefore be viewed as one possible pathway within a broader ongoing effort to develop scientifically grounded AI.

THERMODYNAMIC STABILITY, AI SAFETY, AND THE ROLE OF ZENTROPY

The rapidly increasing complexity of AI models raises a fundamental question: how do we ensure the "mental health" of AI systems as they scale? Just as HI evolved under the constraints of biology, environment, and culture, modern AI is shaped by the vast flows of information fed into its training processes. Recent public discussions, exemplified by Suleyman's The Coming Wave[67], which outlines the destabilizing potential of uncontrolled technological growth, and Diamond's Guns, Germs, and Steel[68], which attributes civilizational trajectories to structural environmental constraints, highlight a common theme: systems remain stable only when their internal dynamics are grounded in robust, interpretable principles. In large-scale neural networks, however, training can easily fall into pathological regimes: excessive tuning parameters may lead to behavior analogous to depression (models collapsing to overly pessimistic or underconfident predictions), local minima may trap the model in non-functional states, and runaway divergence may resemble bipolar instability, oscillating between extreme outputs without the ability to "descend" to a stable solution. These phenomena are not literal mental illnesses, but they metaphorically capture the fragility of high-dimensional optimization landscapes lacking principled structure.

Zentropy theory and the ZENN framework offer a scientifically grounded path toward addressing these concerns. By modeling AI systems as statistical ensembles of configurations governed by physically meaningful energy and entropy relationships, ZENN replaces *ad-hoc* loss landscapes with thermodynamically constrained ones. Cross-zentropy introduces intrinsic entropy into the learning objective, preventing pathological overconfidence and smoothing the optimization landscape in a way analogous to thermodynamic stabilization[2]. The configuration-based architecture of ZENN provides an additional layer of structural regularization: instead of relying solely on raw features or token-level correlations, ZENN learns an ensemble of meaningful internal configurations, each with its own internal energy and intrinsic entropy. This ensemble-based formulation mitigates collapse, reduces susceptibility to extreme oscillatory modes, and provides a potential pathway that the model's macroscopic predictions correspond to well-defined free-energy minima, properties directly relevant to AI robustness and safety.

Consequently, ZENN can be viewed as providing not just better accuracy but improved AI psychological hygiene: a framework in which models remain stable, interpretable, and resistant to pathological optimization dynamics[2]. By embedding thermodynamic principles into learning, ZENN gives AI systems something akin to a scientific internal compass: a structured, physically grounded mechanism for regulating uncertainty, avoiding extremes, and maintaining stable behavior under diverse conditions. In this sense, zentropy theory may ultimately inform not only better-performing AI models but also safer and more reliable ones, offering a foundation for the next generation of science-based, safety-aligned intelligence.

A deeper implication of this perspective comes from correcting a widespread misconception surrounding "ergodicity breaking"[69-73] in complex systems that confuses the ergodic accessibility of configurations and the accessibility of states with the latter being the basins on the free-energy landscape. As clarified in our thermodynamic analysis, metastable states do not arise from any failure of ergodicity but from the structure of the free-energy landscape itself: a thermodynamic state is always a statistical mixture of all symmetry-broken configurations permitted by the ensemble as shown by zentropy theory, i.e., from Equation 1 to Equation 4, even when transitions between states are dynamically or kinetically hindered. This distinction is essential because it grounds stability, metastability, and instability firmly in energy-entropy competition rather than in dynamical assumptions. By properly accounting for the full configuration space, including excited, symmetry-broken, and low-probability configurations, zentropy provides a principled mechanism for predicting the shape of the free-energy landscape and the resulting thermodynamic behavior. This predictive capability underpins why zentropy successfully captures phenomena such as stability, instability, negative thermal expansion, and criticality in physical systems: these behaviors emerge from the geometry and curvature of the free-energy surface, not from any breakdown of statistical mechanics.

This insight extends directly to AI safety. When AI models are framed as statistical ensembles of internal configurations, their macroscopic behavior is likewise governed by an effective free-energy landscape. Zentropy enables this landscape to be computed rather than guessed, providing a quantitative means to identify stable operating regimes, detect approaching instabilities, and forecast critical transitions in model behavior[2]. Such transitions, analogous to phase changes in physical systems, may correspond to sudden failures, runaway amplification, catastrophic misalignment, or collapse into degenerate modes. By modeling the ensemble properly and evaluating its free energy, zentropy supplies an early-warning mechanism grounded in first principles[2]. In this way, the same framework that predicts physical singularities and critical points also establishes a scientific basis for ensuring that AI systems remain stable, interpretable, and safe, even as their architectures and autonomy levels continue to grow.

Furthermore, beyond stabilizing learning dynamics, zentropy theory provides a natural foundation for predicting singularities, critical points, and bifurcations in AI behavior. Because zentropy models the free-energy landscape of an ensemble of configurations, abrupt qualitative changes in the model's output, analogous to phase transitions, correspond to structural features such as non-analyticities, vanishing curvatures, or competing minima in the free-energy surface. This is precisely how physical systems identify critical points, emergent order, or bifurcation pathways. In the same spirit, ZENN can detect when an AI system approaches instability, transitions between competing representational modes, or enters pathological regions of parameter space. As a result, zentropy offers a scientifically interpretable mechanism for forecasting when an AI model may undergo sudden shifts, suggesting the potential for a thermodynamics-inspired framework to identify instability and abrupt transitions in AI systems for robustness and safety[2,74]. By treating the model as an ensemble-based dynamical system rather than a single monolithic function, ZENN enables the prediction, monitoring, and control of these critical phenomena with mathematical rigor.

SUMMARY AND OUTLOOK

In summary, the HI evolution, from neural processing to the layered abstractions of thermodynamics, statistical mechanics and quantum mechanics, provides a powerful blueprint for the next stages of AI evolution. Modern AI architectures echo this progression, yet they remain incomplete until foundational scientific principles are explicitly incorporated into their design. BENN and ZENN advance this vision by embedding the core structures of statistics, quantum mechanics, and statistical mechanics directly into neural networks, yielding models that are not only more capable but also physically interpretable and scientifically principled.

This integration moves AI beyond pattern recognition into a science-based paradigm capable of reasoning across scales: from electronic structure and configurations to macroscopic behavior and emergent phenomena. Crucially, the thermodynamic grounding of ZENN also speaks directly to AI safety and stability. As AI systems grow larger and more autonomous, their optimization landscapes become increasingly complex and susceptible to pathological behaviors such as mode collapse, runaway amplification, brittle decision boundaries, or catastrophic errors triggered by small perturbations. Zentropy theory has the potential to provide a stabilizing mechanism by imposing thermodynamic constraints on learning, ensuring that the model evolves toward physically meaningful free-energy minima rather than arbitrary or unstable solutions. In doing so, ZENN naturally identifies critical points, singularities, and bifurcations in model behavior, features often invisible in conventional architectures but essential for predicting abrupt transitions or safety-critical failure modes.

By structuring AI systems as ensembles of configurations governed by well-defined energy and entropy contributions, ZENN provides a principled framework for maintaining stability, avoiding pathological extremes, and ensuring that system behavior remains interpretable and predictable as models scale. By analogy with human cognition, these activities may be viewed as preserving the "mental health" of AI systems. As AI increasingly becomes a central engine for scientific discovery and technological innovation, such thermodynamically grounded frameworks will be essential not only for improving accuracy, but also for ensuring that AI systems behave safely, robustly, and in accordance with the fundamental principles governing nature. In this sense, BENN and ZENN represent an important step toward transparent, theory-driven, and safety-informed intelligence for the next era of discovery.

DECLARATIONS

Acknowledgments

The authors are grateful to Yin Tang for redrawing Figure 3.

Authors' contributions

Prepared the draft: Liu, Z. K.
Contributed to the revision of the manuscript: Hao, W; Bing Li, B.

Availability of data and materials

Not applicable.

AI and AI-assisted tools Statement

During the preparation of this manuscript, the AI tools Microsoft Copilot and Claude were used for language editing and for generating certain icons used in the graphical abstract. The use of these tools was limited to language refinement and auxiliary visual design. The tools did not influence the study design, data collection, data analysis, interpretation, or the scientific content of the work. All authors take full responsibility for the accuracy, integrity, and final content of the manuscript.

Financial support and sponsorship

None.

Conflicts of interest

Liu, Z. K. is Editor-in-Chief of the journal *ZENtropy*. Hao, W. is a Deputy Editor-in-Chief of *ZENtropy*. None of them were involved in any part of the editorial process, including reviewer selection, manuscript handling, or decision-making. The other author declare that there are no conflicts of interest.

Ethical approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

REFERENCES

1. Tang, Y.; Li, B. Belted and ensembled neural network for linear and nonlinear sufficient dimension reduction. *J. Am. Stat. Assoc.* **2026**, 1-12. DOI
2. Wang, S.; Shang, S.; Liu, Z.; Hao, W. ZENN: A thermodynamics-inspired computational framework for heterogeneous data-driven modeling. *Proc. Natl. Acad. Sci. U.S.A.* **2026**, *123*, e2511227122. DOI

3. Azevedo, F. A.; Carvalho, L. R.; Grinberg, L. T.; et al. Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain. *J. Comp. Neurol.* **2009**, *513*, 532-41. DOI
4. Lent, R.; Azevedo, F. A. C.; Andrade-Moraes, C. H.; Pinto, A. V. O. How many neurons do you have? Some dogmas of quantitative neuroscience under revision. *Eur. J. Neurosci.* **2011**, *35*, 1-9. DOI
5. Kande, E. R.; Koester, J. D.; Mack, S. H.; Siegelbaum, S. A. Principles of neural science, 6th ed.; McGraw-Hill, 2021.
6. Hohenberg, P.; Kohn, W. Inhomogeneous electron gas. *Phys. Rev.* **1964**, *136*, B864-71. DOI
7. Kohn, W.; Sham, L. J. Self-consistent equations including exchange and correlation effects. *Phys. Rev.* **1965**, *140*, A1133-8. DOI
8. Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **1996**, *77*, 3865-8. DOI
9. Liu, Z.; Wang, Y.; Shang, S. Zentropy theory for positive and negative thermal expansion. *J. Phase. Equilib. Diffus.* **2022**, *43*, 598-605. DOI
10. Wang, Y.; Hector, L. G.; Zhang, H.; Shang, S. L.; Chen, L. Q.; Liu, Z. K. Thermodynamics of the Ceγ-αtransition: density-functional study. *Phys. Rev. B.* **2008**, *78*, 104113. DOI
11. Wang, Y.; Hector Jr, L. G.; Zhang, H.; Shang, S. L.; Chen, L. Q.; Liu, Z. K. A thermodynamic framework for a system with itinerant-electron magnetism. *J. Phys:. Condens. Matter.* **2009**, *21*, 326003. DOI
12. Wang, Y.; Shang, S.; Zhang, H.; Chen, L.; Liu, Z. Thermodynamic fluctuations in magnetic states: Fe_3Pt as a prototype. *Philos. Mag. Lett.* **2010**, *90*, 851-9. DOI
13. Perdew, J. P. SCAN meta-GGA, strong correlation, symmetry breaking, self-interaction correction, and semi-classical limit in density functional theory: Hidden connections and beneficial synergies? *APL. Computational. Physics.* **2025**, *1*, 010903. DOI
14. Minsky, M., Papert, S. A. *Perceptrons: an introduction to computational geometry*. The MIT Press, 2017. DOI
15. Rumelhart, D. E.; Hinton, G. E.; Williams, R. J. Learning representations by back-propagating errors. *Nature* **1986**, *323*, 533-6. DOI
16. Lecun, Y.; Boser, B.; Denker, J. S.; et al. Backpropagation applied to handwritten zip code recognition. *Neural. Comput.* **1989**, *1*, 541-51. DOI
17. Lecun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE.* **1998**, *86*, 2278-324. DOI
18. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural. Comput.* **1997**, *9*, 1735-80. DOI
19. Scarselli, F.; Gori, M.; Tsoi, A. C.; Hagenbuchner, M.; Monfardini, G. The graph neural network model. *IEEE. Trans. Neural. Netw.* **2009**, *20*, 61-80. DOI
20. Vaswani, A.; Brain, G.; Shazeer, N.; et al. Attention is all you need. arXiv 2017;arXiv:1706.03762. Available online: https://arxiv.org/abs/1706.03762. [accessed 9 July 2026].
21. Radford, A., Narasimhan, K., Salimans, T. & Sutskever, I. Improving language understanding by generative pre-training. 2018. Available from: https://www.mikecaptain.com/resources/pdf/GPT-1.pdf.
22. Raissi, M.; Perdikaris, P.; Karniadakis, G. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J. Comput. Phys.* **2019**, *378*, 686-707. DOI
23. Lewis, P.; Perez, E.; Piktus, A.; et al. Retrieval-augmented generation for knowledge-intensive NLP tasks. *Adv. Neural. Inf. Process. Syst.* **2021**, *33*, 9459-74. https://proceedings.neurips.cc/paper/2020/hash/6b493230205f780e1bc26945df7481e5-Abstract.html.
24. Cobbe, K.; Kosaraju, V.; Bavarian, M.; et al. Training verifiers to solve math word problems. arXiv 2021;arXiv:2110.14168. Available online: http://arxiv.org/abs/2110.14168. [accessed 9 July 2026].
25. Ouyang, L.; Wu, J.; Jiang, X.; et al. Training language models to follow instructions with human feedback. **2022**, *35*, 27730-44. DOI
26. Wei, J.; Wang, X.; Schuurmans, D.; et al. Chain-of-thought prompting elicits reasoning in large language models. *Adv. Neural. Inf. Process. Syst.* **2023**, *35*, 24824-37. DOI
27. Schick, T.; Dwivedi-Yu, J.; Dessì, R.; et al. Toolformer: language models can teach themselves to use tools. *Adv. Neural Inf. Process. Syst.* 2023, *36*, 68539-51. DOI
28. Huang, Y.; Hao, W.; Lin, G. HomPINNs: Homotopy physics-informed neural networks for learning multiple solutions of nonlinear elliptic differential equations. *Comput. Math. Appl.* **2022**, *121*, 62-73. DOI
29. Zheng, H.; Huang, Y.; Huang, Z.; Hao, W.; Lin, G. HomPINNs: homotopy physics-informed neural networks for solving the inverse problems of nonlinear differential equations with multiple solutions. *J. Comput. Phys.* **2024**, *500*, 112751. DOI PubMed PMC
30. E, W.; Yu, B. The Deep Ritz Method: A deep learning-based numerical algorithm for solving variational problems. *Commun. Math. Stat.* **2018**, *6*, 1-12. DOI
31. Lu, L.; Jin, P.; Pang, G.; Zhang, Z.; Karniadakis, G. E. Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators. *Nat. Mach. Intell.* **2021**, *3*, 218-29. DOI
32. Li, Z.; Huang, D. Z.; Liu, B.; Anandkumar, A. Fourier neural operator with learned deformations for PDEs on general geometries. *J. Mach. Learn. Res.* **2023**, *24*, 1-26. http://jmlr.org/papers/v24/23-0064.html.

33. Hao, W.; Liu, X.; Yang, Y. Newton informed neural operator for solving nonlinear partial differential equations. *Adv. Neural. Inf. Process. Syst.* **2024**, *37*, 120832. PMC11973962. DOI
34. Wang, J.; Hao, W. Laplacian eigenfunction-based neural operator for learning nonlinear reaction-diffusion dynamics. *J. Comput. Phys.* **2025**, *543*, 114400. DOI PubMed PMC
35. Clausius, R. Ueber verschiedene für die Anwendung bequeme Formen der Hauptgleichungen der mechanischen Wärmetheorie. *Annalen. der. Physik.* **2006**, *201*, 353-400. (in German). DOI
36. Boltzmann, L. Weitere Studien über das Wärmegleichgewicht unter Gasmolekülen. Sitzungsberichte der Kais. *Akad. der. Wissenschaften.* **1872**, *66*, 275-370. (in German). DOI
37. Gibbs, J. W. Elementary principles in statistical mechanics: developed with especial reference to the rational foundation of thermodynamics; Cambridge University Press, 2011. DOI
38. Shannon, C. E. A mathematical theory of communication. *Bell. System. Technical. Journal.* **1948**, *27*, 379-423. DOI
39. Kullback, S.; Leibler, R. A. On Information and Sufficiency. *Ann. Math. Statist.* **1951**, *22*, 79-86. DOI
40. Jaynes, E. T. Information theory and statistical mechanics. *Phys. Rev.* **1957**, *106*, 620-30. DOI
41. Bridle, J. S. Probabilistic Interpretation of Feedforward Classification Network Outputs, with Relationships to Statistical Pattern Recognition. In *Neurocomputing*; Soulié, F. F., Hérault, J., Eds.; Springer Berlin Heidelberg, 1990; pp 227-36. DOI
42. Richard, M. D.; Lippmann, R. P. Neural network classifiers estimate bayesian *a posteriori* probabilities. *Neural. Comput.* **1991**, *3*, 461-83. DOI
43. Quinlan, J. R. Induction of decision trees. *Mach. Learn.* **1986**, *1*, 81-106. DOI
44. Wainwright, M. J.; Jordan, M. I. Graphical Models, Exponential Families, and Variational Inference. *Found. Trends. Mach. Learn.* **2008**, *1*, 1-305. DOI
45. Kingma, D. P.; Welling, M. Auto-encoding variational bayes. arXiv 2014;arXiv:1312.6114. Available online: http://arxiv.org/abs/1312.6114. [accessed 9 July 2026].
46. Haarnoja, T.; Zhou, A.; Abbeel, P.; Levine, S. Soft actor-critic: off-policy maximum entropy deep reinforcement learning with a stochastic actor. arXiv 2018;arXiv:1801.01290. Available online: http://arxiv.org/abs/1801.01290. [accessed 9 July 2026].
47. Holtzman, A.; Buys, J.; Du, L.; Forbes, M.; Choi, Y. The curious case of neural text degeneration. arXiv 2020;arXiv:1904.09751. Available online: http://arxiv.org/abs/1904.09751. [accessed 9 July 2026].
48. Bayes, T. LII. An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, F. R. S. communicated by Mr. Price, in a letter to John Canton, A. M. F. R. S. *Phil. Trans. R. Soc. A.* **1763**, 370-418. DOI
49. Myers, L. A.; Hew, N. L. E.; Shang, S. L.; Liu, Z. K. Recursive entropy in thermodynamics: establishing the statistical-physics basis of the zentropy approach. arXiv 2025;arXiv:2511.04950. Available online: http://arxiv.org/abs/2511.04950. [accessed 9 July 2026].
50. Wang, Y.; Liu, Z.; Chen, L. Thermodynamic properties of Al, Ni, NiAl, and Ni_3Al from first-principles calculations. *Acta. Mater.* **2004**, *52*, 2665-71. DOI
51. Liu, Z.; Li, B.; Lin, H. Multiscale entropy and its implications to critical phenomena, emergent behaviors, and information. *J. Phase. Equilib. Diffus.* **2019**, *40*, 508-21. DOI
52. Gibbs, J. W. The collected works of J. Willard Gibbs: Vol. II Elementary Principles in Statistical Mechanics. Yale University Press, **1948**, Vol. II.
53. Liu, Z. Computational thermodynamics and its applications. *Acta. Mater.* **2020**, *200*, 745-92. DOI
54. Liu, Z. Thermodynamics and its prediction and CALPHAD modeling: review, state of the art, and perspectives. *Calphad* **2023**, *82*, 102580. DOI
55. Liu, Z. Quantitative predictive theories through integrating quantum, statistical, equilibrium, and nonequilibrium thermodynamics. *J. Phys. Condens. Matter.* **2024**, *36*, 343003. DOI
56. Wang, S.; Shang, S. L.; Liu, Z. K.; Hao, W. ZENN on GitHub. **2026**. https://github.com/WilliamMoriaty/ZENN. [accessed 9 July 2026].
57. Liu, Z.; Hew, N. L. E.; Shang, S. Zentropy theory for accurate prediction of free energy, volume, and thermal expansion without fitting parameters. *Microstructures* **2024**, *4*, 2024009. DOI
58. Yuan, K.; Miao, D.; Yao, Y.; Zhang, H.; Zhao, X. Feature selection using zentropy-based uncertainty measure. *IEEE. Trans. Fuzzy. Syst.* **2024**, *32*, 2246-60. DOI
59. Yuan, K.; Miao, D.; Pedrycz, W.; Ding, W.; Zhang, H. Ze-HFS: zentropy-based uncertainty measure for heterogeneous feature selection and knowledge discovery. *IEEE. Trans. Knowl. Data. Eng.* **2024**, *36*, 7326-39. DOI
60. Yuan, K.; Miao, D.; Pedrycz, W.; Zhang, H.; Hu, L. Multigranularity data analysis with zentropy uncertainty measure for efficient and robust feature selection. *IEEE. Trans. Cybern.* **2025**, *55*, 740-52. DOI
61. Yuan, K.; Miao, D.; Zhang, H.; Pedrycz, W. An efficient and robust feature selection approach based on zentropy measure and neighborhood-aware model. *IEEE. Trans. Neural. Netw. Learning. Syst.* **2025**, *36*, 16351-65. DOI

62. Ye, W.; Xu, W. Innovative multi-granularity granular-balls rough set for feature selection: Driving generalized multi-granularity rough set evolution with Zentropy integration. *Inform. Sciences.* **2025**, *718*, 122411. DOI
63. Dong, H.; Liu, C.; Chen, X.; Miao, D. A multi-granularity decision tree algorithm based on variable precision rough sets and Zentropy. *Appl. Soft. Comput.* **2025**, *185*, 113851. DOI
64. Yuan, K.; Miao, D.; Ding, W.; Pedrycz, W.; Yao, Y. Robust semi-supervised feature selection with multi-granularity zentropy modeling. *IEEE. Trans. Pattern. Anal. Mach. Intell.* **2026**, *48*, 4587-604. DOI
65. Einstein, A. Autobiographical notes. Albert Einstein: philosopher-scientist. Library of Living Philosophers, **1949**. https://www.amazon.com/Albert-Einstein-Philosopher-Scientist-Philosophers-Paperback/dp/0875482864.
66. Brush, S. G. The kind of motion we call heat. North-Holland, 1976.
67. Suleyman, M.; Bhaskar, M. The coming wave: technology, power, and the twenty-first century's greatest dilemma. Crown, an imprint of Crown Publishing Group, 2023.
68. Diamond, J. M. Guns, germs, and steel: the fates of human societies. W. W. Norton & Company, 1999.
69. Palmer, R. Broken ergodicity. *Adv. Phys.* **2006**, *31*, 669-735. DOI
70. Binder, K.; Young, A. P. Spin glasses: experimental facts, theoretical concepts, and open questions. *Rev. Mod. Phys.* **1986**, *58*, 801-976. DOI
71. Bouchaud, J. P. Weak ergodicity breaking and aging in disordered systems. *J. Phys. I. France.* **1992**, *2*, 1705-13. DOI
72. Cugliandolo, L. F.; Kurchan, J. Analytical solution of the off-equilibrium dynamics of a long-range spin-glass model. *Phys. Rev. Lett.* **1993**, *71*, 173-6. DOI
73. Berthier, L.; Biroli, G. Theoretical perspective on the glass transition and amorphous materials. *Rev. Mod. Phys.* **2011**, *83*, 587-645. DOI
74. Liu, Z. A Unified Thermodynamic framework: from equilibrium and nonequilibrium to zentropy, cross phenomena, and applications for AI and AI safety. *J. Phase. Equilib. Diffus.* **2026**, 1248. DOI