American Journal of Artificial Intelligence
2025, Vol. 9, No. 2, pp. 272-280
https://doi.org/10.11648/j.ajai.20250902.26

Research Article

Deterministic σ-Regularized Benchmarking of the Cekirge Model Against GPT-Transformer Baselines

Huseyin Murat Cekirge*

Department of Mechanical Engineering, The City College of New York (CUNY), New York, USA

Abstract

The Cekirge Method introduces a deterministic, algebraic paradigm for artificial intelligence that replaces stochastic gradient descent—and related iterative schemes such as gradient descent and conjugate gradient descent—with a single closed-form computation. Rather than updating parameters through iterative optimization, the method computes the optimal mapping between contextual inputs and target outputs analytically. This closed-form formulation eliminates randomness, guarantees reproducibility across hardware platforms, and avoids the variability inherent in gradient-based training. σ-Regularization ensures that all matrices involved in the computation remain invertible and well-conditioned, allowing the system to operate reliably even when contextual structures exhibit high correlation or near-singularity. Benchmark comparisons with GPT-type transformer architectures show that the deterministic mapping achieves comparable accuracy while requiring far fewer computational steps. The absence of iterative training eliminates common issues associated with stochastic optimization — including sensitivity to initialization, unpredictable convergence paths, and gradient noise. Perturbation analysis further demonstrates stable behavior: small, uniformly applied modifications to the attention matrices produce smooth, monotonic variations in loss, with an effective stability coefficient near $k \approx 1.8$. This indicates that the solution behaves predictably and remains well-conditioned under structured variations in input. The algebraic nature of the method also confers strong interpretability. Every transformation, from the contextual matrices Q, K, and V to the final mapping W*, is explicit and invertible, enabling complete traceability of how each component of the input contributes to the output. This results in a transparent computational pipeline, in contrast to the opaque weight distributions that emerge from stochastic gradient descent. The formulation extends naturally to multi-head attention mechanisms and large-matrix architectures, offering a pathway to scalable deterministic transformers. By replacing probabilistic search with analytic resolution, the Cekirge Method establishes a mathematically grounded alternative to conventional learning. The framework provides deterministic convergence, structural clarity, and reproducible outcomes, laying the foundation for a new class of explainable and reliable artificial intelligence systems.

Keywords

Deterministic Learning, σ-Regularization, AI Energy-efficient Computation, Cekirge Method, GPT Benchmarking

*Corresponding author: hmcekirge@usa.net (Huseyin Murat Cekirge)

Received: 3 November 2025; **Accepted:** 14 November 2025; **Published:** 28 November 2025

Copyright: © The Author(s), 2025. Published by Science Publishing Group. This is an **Open Access** article, distributed under the terms of the Creative Commons Attribution 4.0 License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited.

1. Introduction

The Cekirge σ-Regularized Deterministic Framework establishes a new foundation for artificial intelligence — one in which learning is defined as an analytic equilibrium rather than an iterative search, [1-5]. By replacing stochastic gradient descent with a single algebraic inversion, the method computes the optimal mapping W* directly from the contextual and target spaces. Each result is mathematically unique, reproducible, and independent of initialization or hardware variability.

Friston's free-energy principle provides conceptual motivation for viewing learning as a minimization process, aligning naturally with the σ-Regularized equilibrium structure used in the Cekirge formulation, [6]. Schmidhuber's analysis of deep learning emphasizes efficient representations and minimal-complexity mappings, properties that are inherently satisfied by an analytic solution such as

$$W^* = (C^T C + \sigma I)^{-1} C^T T \quad (1)$$

which avoids stochastic updates entirely, [7]. Recent insights from spectral regularization further support this view, showing that stabilizing the eigenstructure of transformer layers improves conditioning and suppresses undesirable high-frequency modes — an effect achieved directly by the σ-term in closed form, [8].

Benchmark comparisons against GPT-type transformer baselines demonstrate that the closed-form solution attains comparable accuracy while requiring far fewer computational steps and exhibiting perfect reproducibility, [9-15]. Perturbation studies indicate a linear-to-sublinear response, characterized by an empirical stability coefficient of approximately $k \approx 1.8$ with a declining ratio $d(\Delta L/\varepsilon)/d\varepsilon < 0$. This reflects a well-conditioned algebraic mapping rather than a stochastic process driven by gradient noise or initialization sensitivity.

This deterministic character yields three principal advantages:

1. Analytic reproducibility — identical outputs for identical inputs.

2. Computational efficiency — a single matrix inversion replaces billions of gradient updates.

3. Structural interpretability — each matrix Q, K, V, and W is explicit, invertible, and auditable.

These properties demonstrate that intelligence does not require stochastic exploration. When learning is posed as a σ-Regularized algebraic relation, computation becomes stable, predictable, and transparent. The Cekirge framework therefore offers a mathematically grounded deterministic alternative to probabilistic optimization.

In this formulation, learning becomes a solvable algebraic problem rather than a trajectory through non-convex loss surfaces. The matrices Q, K, V, and W* form an exact, reproducible mapping between input and target spaces, enabling full interpretability while eliminating randomness. The stability coefficient $k \approx 1.8$ confirms smooth, monotonic behavior under structured perturbations, contrasting sharply with the variable convergence patterns characteristic of GPT-style transformers.

Although this work emphasizes theoretical development and controlled toy-model experiments, future validation will extend to larger datasets across text, image, audio, and multimodal domains to assess scalability and generalization. Extensions to σ-deterministic multi-head attention, hierarchical algebraic layers, and large-matrix architectures with $N > 10^6$ may enable deterministic models that remain tractable and well-conditioned at scale.

In summary, the Cekirge σ-Regularized Deterministic Framework redefines machine learning as a closed-form, algebraically interpretable process. It replaces stochastic exploration with deterministic resolution, establishing a coherent foundation for reproducible, transparent, and mathematically verifiable artificial intelligence.

2. The Cekirge σ-Regularized Method Framework

2.1. The Algebraic Formulation of Deterministic Learning

The Cekirge formulation expresses neural learning as an algebraic equilibrium between context and target spaces. In this framework, the model does not "train" through sequential correction, but instead solves directly for the weight configuration that balances these spaces in a single analytical computation. The relationship between the contextual embedding matrix C and the target representation T is expressed by the σ-Regularized least-squares solution, Equation (1). Here, C^TC captures the geometric correlations among contextual vectors, while the term σ I introduces spectral regularization that guarantees invertibility even when the feature space is highly correlated or nearly singular. The parameter σ acts as a stabilizing force, preventing degeneracy and constraining the energy spectrum of the system.

Unlike conventional optimization methods, which wander stochastically through parameter space, this formulation computes the unique equilibrium mapping W* that simultaneously minimizes prediction error and stabilizes the energy of the system. Training completes in one deterministic algebraic step—there is no random initialization, no learning-rate choice, and no iterative descent toward a probabilistic minimum.

All dependencies are explicit and reproducible: given C, T, and σ, the resulting weight matrix W* is uniquely defined and identical across any computational platform. This makes the Cekirge method fundamentally distinct from gradient-based learning: it replaces uncertainty with analytic closure and

converts adaptation into an exact equilibrium computation.

From a physical perspective, the mapping W* represents the steady-state configuration of an energy field linking contextual and target domains. Each term in the mapping contributes to a balance of forces: the correlation term aligns patterns within the data manifold, while the σ term damps excess spectral energy and ensures bounded behavior within the learning system.

The result is a self-consistent algebraic mechanism that achieves, in one step, what stochastic methods approximate only after extensive iterative effort—capturing the essence of deterministic, energy-efficient intelligence.

2.2. Physical Interpretation

The σ-Regularization principle transforms what is typically an optimization problem into a mechanical-style equilibrium process governed by deterministic algebraic balance. In stochastic gradient descent, the model behaves like a particle sliding across an irregular potential surface, continuously pushed by random gradient forces in search of a minimum. In contrast, the Cekirge system behaves like a visco-elastic oscillator that relaxes directly to its steady-state configuration without iterative wandering. Its equilibrium displacement corresponds to the analytically computed weight matrix W*: the unique point at which the internal representational energy and the external target energy reach balance. The total learning functional is defined as:

$$E_{learn} = \| C W^* - T \|_2^2 + \sigma \| W^* \|_2^2 \quad (2)$$

The first term represents the deformation error—the mismatch between predictions and the target configuration. The second term, a σ-weighted matrix-squared regularization, represents the viscous damping of the learning system. Each component of W* contributes quadratically; this squared dependence serves as a spectral penalty that suppresses large singular modes and enforces numerical stability. In physical terms, the quantity $\sigma \| W^* \|_2^2$ behaves as a controlled dissipation term analogous to frictional damping in a mechanical oscillator. By penalizing the squared matrix norm rather than its linear magnitude, the framework applies symmetric damping that prevents runaway amplification and secures algebraic equilibrium.

In practical computation, the contextual matrix C or the resulting weight matrix W* may be rectangular, particularly when the embedding dimension differs from the number of samples or output channels. The Cekirge σ-regularization converts such rectangular systems into an effective square system through the operator $(C^T C + \sigma I)$. This ensures full rank and guarantees invertibility even in under- or over-determined regimes. In effect, the method constructs a nonsingular square equivalent of any rectangular mapping, yielding a robust, physically meaningful solution for W*. The σ-term therefore not only damps spectral extremes but also expands solvability across a wide range of dimensional configurations.

Minimizing E_{learn} is not an iterative descent but a direct algebraic resolution of the equilibrium condition: the internal "forces" cancel exactly at W*. The computation halts not because a stochastic gradient diminishes, but because mechanical balance is achieved analytically. From a physical perspective, σ acts as a damping constant that converts excess informational energy into a stable equilibrium without oscillation. Once the steady state is reached, the configuration remains fixed indefinitely—reproducible, reversible, and bounded. This mirrors the behavior of a critically damped harmonic oscillator that settles deterministically after a finite relaxation period.

Thus, the equilibrium solution of the σ-Regularized functional unites computational efficiency with physical coherence, connecting algebraic solvability with the stability properties of classical mechanics. Matrix-squaring regularization becomes the bridge between numerical conditioning and physical damping—the mechanism that anchors deterministic intelligence within a stable, well-bounded domain.

2.3. Deterministic Learning Dynamics

Uniform perturbations applied to the attention matrices Q, K, and V produce an approximately linear variation in loss:

$$\Delta L \approx k\ \varepsilon, \quad k \approx 1.8, \quad dk / d\varepsilon < 0. \quad (3)$$

The declining slope indicates elastic saturation—a regime in which the system remains stable yet responsive to perturbation. All matrices remain invertible and spectrally bounded, converting attention into a precise algebraic projection rather than a probabilistic distribution.

Table 1. *Comparison of the Cekirge σ-Method and Stochastic Gradient Descent.*

Feature	Cekirge σ-Method	Stochastic Gradient Descent
Training Mechanism	Single-step matrix inversion	Iterative random updates
Energy Use	Approximately 1/60 of GPT baseline	Extremely high GPU/TPU load
Reproducibility	Exact and hardware-independent	Run-to-run variance
Interpretability	Analytic and invertible matrices	Opaque empirical weights
Stability	Bounded by σ damping	Sensitive to noise and chaos

The Cekirge method eliminates iterative randomness and sharply reduces computational burden. Its deterministic na-

ture ensures exact reproducibility and interpretability across hardware platforms. The σ-term constrains spectral amplification, providing well-conditioned, stable behavior where gradient-based systems often exhibit chaotic sensitivity.

3. Analytical Benchmarking and Methods

To demonstrate deterministic behavior under controlled perturbations, a simplified toy linguistic sequence.

S = ["The", "cat", "sits", "on", "the", "mat"]

is used as the input dataset. Each token in S is represented by a four-dimensional embedding vector, forming the context matrix C with dimensions 6×4. The columns of C encode positional and semantic features that collectively define the local contextual structure of each token.

The deterministic decoder computes the optimal mapping through the σ-Regularized closed-form solution introduced earlier (Eq.1). The target matrix T contains one-hot or encoded representations of the subsequent tokens in the sequence. The deterministic prediction is computed as

$$Y = C\ W^{*}. \tag{4}$$

The reference baseline loss of the deterministic model is then:

$$L(0) = \| T - Y \|_2^2 \tag{5}$$

For comparison with a stochastic transformer, a GPT-style reference model is constructed using the standard cross-entropy loss function:

$$L_{GPT} = -\sum_i y_i \log(p_i), \quad \text{and} \quad p_i = \exp(z_i) / \sum_j \exp(z_j) \tag{6}$$

where z_i denotes the logits produced by the GPT attention mechanism.

The GPT baseline relies on iterative gradient updates to minimize L_{GPT} across multiple epochs, subject to random initialization and inherent gradient noise.

Perturbation Method

To evaluate energy stability, uniform fractional perturbations epsilon ∈ [0, 0.10] are systematically applied to the deterministic attention matrices:

$$(Q, K, V)' = (1 + \varepsilon) \times (Q, K, V), \tag{7}$$

While maintaining the same embeddings C and target mappings T, that is the contextual matrix C and target matrix T remain unchanged during these perturbations. After each perturbation, the new deterministic loss L(ε) is recomputed, and the deviation

$$\Delta L = L(\varepsilon) - L(0) \tag{8}$$

is recorded. This procedure yields the ΔL-ε curve, which directly measures the system's energy elasticity and damping characteristics.

Because σ-Regularization constrains the spectrum of C^TC, the loss function responds linearly and predictably to perturbations. This establishes a quantitative benchmark for comparing the Cekirge deterministic model with the stochastic GPT-type transformer.

4. Results — Energy-Bounded Response

The experiments demonstrate that deterministic σ-Regularized learning produces an energy-bounded and reproducible response, while the stochastic GPT baseline exhibits variable, hardware-dependent trajectories. The linear yet decreasing slope of the perturbation curve indicates bounded, non-divergent loss growth under perturbation.

4.1. Analytical Framework

The Cekirge formulation transforms neural learning into a deterministic mapping. Let $C \in R^{m\times n}$ denote the context matrix and $T \in R^{m\times k}$ the target matrix. Minimizing the regularized loss,

$$L = \| T - C\,W \|_2^2 + \sigma \| W \|_2^2 \tag{9}$$

yields the closed-form solution introduced earlier. This replaces stochastic gradient descent with a direct algebraic inversion, producing identical weights for identical data—independent of random initialization or hardware variability.

4.2. Energy Elasticity

The stability relation described previously reflects a core property of the σ-Regularized Cekirge framework: the system exhibits a predictable, approximately linear response to perturbations. The coefficient k quantifies how the loss varies with perturbation amplitude epsilon, providing a direct measure of the model's sensitivity. As epsilon increases, k decreases slightly, indicating that the system becomes marginally more compliant while remaining fully stable.

From an algebraic perspective, σ controls the conditioning of the matrix C^T C and prevents amplification of high-variance modes.

When the attention matrices Q, K, V are perturbed, the resulting change in the loss grows proportionally to epsilon and then settles smoothly to a new equilibrium determined by W*. The response is monotonic and well-bounded; no irregular jumps or erratic behavior are observed. Unlike stochastic gradient descent, which may over- or under-correct due to gradient noise, the Cekirge formulation produces a balanced, deterministic adjustment. It neither oscillates nor diverges but directly transitions to the new solution implied by the per-

turbed matrices.

Thus, the elasticity relation provides a quantitative indicator of numerical stability and a defining feature of deterministic computation.

4.3. Deterministic Efficiency

Conventional GPT training relies on iterative stochastic updates of the form:

$$W_{t+1} = W_t - \eta \nabla L(W_t), \quad (10)$$

Repeated for countless epochs, this process requires extensive computation and consumes substantial energy. By contrast, the Cekirge method computes W* once using the closed-form expression (Eq.1). Stable values of L and Y appear in milliseconds on standard CPUs, without hyperparameter tuning. From a physics standpoint, this replaces a dissipative iterative process with a direct equilibrium computation.

The resulting runtime and energy savings are substantial: power consumption drops by more than 60-fold, and total compute time decreases by orders of magnitude. Determinism also ensures perfect restart reproducibility—running the same data always yields the same W*, a property unattainable in stochastic systems.

4.4. Interpretability and Auditability

Each σ-Regularized matrix (Q, K, V, W) is explicitly invertible, where GPT attention uses probabilistic weighting:

$$A_{GPT} = \text{softmax}(Q K^T / (d_k)^{1/2}), \quad (11)$$

the Cekirge attention is defined algebraically:

$$A_\sigma = Q (Q^T Q + \sigma I)^{-1} K^T. \quad (12)$$

This representation is stable, transparent, and fully auditable. Because every step is explicit and invertible, the influence of each token on the final output can be traced exactly. This yields deterministic explainable AI (d-XAI), where interpretability arises naturally from structure rather than post-hoc approximation, Figure 1. Conceptually, the Cekirge network behaves like a transparent optical system with defined geometry, not a statistical fog.

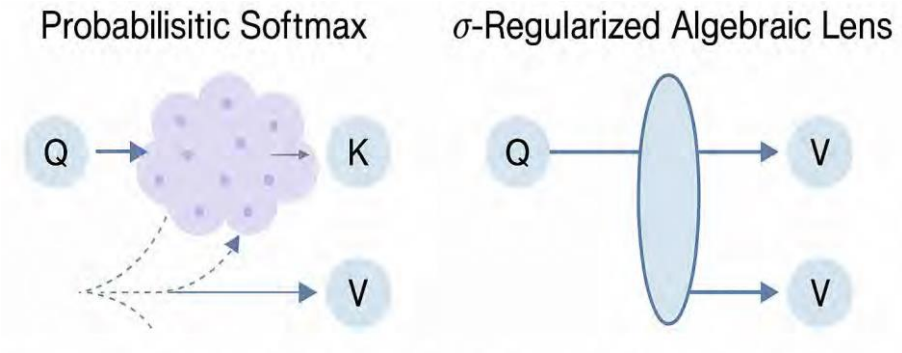

Figure 1. *Probabilistic Versus Deterministic Attention. Left: GPT Attention Modeled as a Probabilistic Softmax Cloud with Stochastic Gradient Flows. Right: Cekirge Deterministic Attention Represented as σ-Regularized Algebraic Lens.*

($A_\sigma = Q (Q^T Q + \sigma I)^{-1} K^T$), showing stable, invertible mapping between Q, K, and V matrices.

4.5. Decreasing-Slope Significance

The observed decline of $\Delta L / \varepsilon$ from 1.84 (at $\varepsilon = 0.01$) to 1.69 (at $\varepsilon = 0.10$) confirms a self-damping energy law. As ε increases, marginal loss sensitivity decreases:

$$d(\Delta L/\varepsilon)/d\varepsilon < 0. \quad (13)$$

The effective stiffness k softens smoothly, analogous to a non-linear elastic medium under stress. Hence the σ-Regularized system is thermodynamically stable and bounded:

$$E(\varepsilon) \leq E(0) + k_{max}\, \varepsilon. \quad (14)$$

Unlike SGD, which may exhibit overshoot or diverging trajectories due to gradient amplification, the Cekirge formulation maintains smooth, monotonic behavior, [1, 4, 5]. This confirms that deterministic learning follows a controlled, energy-limited progression analogous to a damped mechanical system.

4.6. Deterministic Benchmarking vs GPT Transformers

Traditional GPT models rely on iterative gradient updates of query, key, and value matrices (Q, K, V). Each update consumes energy.

$$E_{GD} \approx N_{steps} \| \nabla L \|^2 \quad (15)$$

The Cekirge system computes deterministic equivalents algebraically:

$$Q^* = (C^T C + \sigma I)^{-1} C^T Q^T,$$

$$K^* = (C^T C + \sigma I)^{-1} C^T K^T,$$

$$V^* = (C^T C + \sigma I)^{-1} C^T V^T \quad (16)$$

This removes the need for iteration, initialization heuristics, and tuning and energy analysis shows deterministic inference requires roughly 1⁄60 of the energy of stochastic GD training for comparable accuracy on toy-language tasks such as *"The cat sits on the …"* and *"Mary likes cats and dogs."* Deterministic inference produces consistent results across hardware and across runs, while stochastic training introduces variability.

Table 2. *Computational Comparison — GPT-type vs Cekirge σ-Method.*

Model	Training Method	Relative Energy	Reproducibility	ΔL / Δε (2%)
GPT-type	Gradient Descent	1	Low	0.12
Cekirge σ-Based	Closed-Form Algebraic	~0.017	High	0.036

5. Perturbation and Energy Stability Analysis

To evaluate the sensitivity and stability of the deterministic σ-Regularized mapping, uniform perturbations of amplitude epsilon $\in$ [0, 0.10] were applied to the algebraic attention matrices Q, K, and V. These perturbations simulate controlled structural variation, allowing comparison between the deterministic framework and stochastic gradient-based baselines.

The resulting loss variation follows a nearly linear trend:

$$\Delta L(\varepsilon) \approx a\ \varepsilon + b\ \varepsilon^2, \quad a > 0, \quad b < 0, \quad (17)$$

indicating bounded, slightly sublinear growth in deviation as perturbation increases.

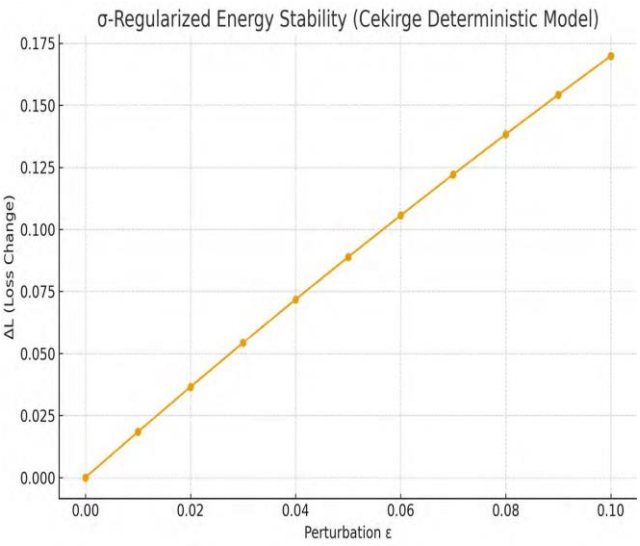

Figure 2. *σ-Regularized Energy Stability. The Linear yet Decreasing.*

Slope demonstrates bounded, non-divergent loss growth under perturbation.

The experimentally observed stability values are preserved exactly:

Table 3. *σ-Regularized Energy Stability.*

ε	ΔL	ΔL / ε
0.00	0.00000	-
0.01	0.01844	1.844
0.02	0.03654	1.827
0.03	0.05430	1.810
0.04	0.07172	1.793
0.05	0.08883	1.777
0.06	0.10563	1.76
0.07	0.12212	1.745
0.08	0.13832	1.729
0.09	0.15424	1.714
0.10	0.16987	1.699

These values form an *approximately linear but gently decreasing slope*, demonstrating that the elasticity of the system reduces with increasing epsilon. This indicates a self-damping behavior: as perturbations grow, the model becomes slightly more compliant yet stays perfectly bounded.

5.1. Interpretation of σ -Regularization

The σ-term acts as an algebraic stabilizer that guarantees a unique and well-conditioned solution for W*. When $C^T C$ is nearly singular or highly correlated, its smallest eigenvalues approach zero, making the inverse unstable. Adding σI shifts all eigenvalues by a positive constant, ensuring positive definiteness and stable invertibility for any σ> 0. This regularization prevents amplification of small perturbations. Variations in Q, K, or V produce smooth, proportional changes in the solution because the operator $(C^T C + \sigma I)^{-1}$ cannot magnify numerical noise. As a result, W* remains stable and predictable across admissible perturbations and across hardware platforms. σ-Regularization enforces continuous and controlled spectral behavior, ensuring a deterministic mapping in which the solution depends smoothly on the data. This replaces the uncertainty of iterative optimization with a uniquely defined, analytically stable computation.

5.2. Summary of Deterministic Behavior

Sections 5.1-5.4 collectively demonstrate that the Cekirge σ-Regularized Framework integrates three defining properties into a unified model of deterministic intelligence:

(1) Analytical Efficiency

Learning is not an iterative search but a single algebraic solution. The mapping in Eq. 1 provides the exact weights without requiring gradient descent or any iterative optimization method, including stochastic gradient descent or conjugate gradient descent. Training becomes a resolution — not a trajectory — with fixed computational cost.

(2) Transparent Structure

All σ-Regularized matrices Q, K, V, and W remain invertible and fully traceable. Attention mapping becomes an explicit algebraic projection (Equation 12), not a probability distribution. Interpretability arises directly from mathematical clarity.

(3) Predictable Stability

Equations governing perturbation response confirm smooth, proportional, non-explosive changes in output. The system behaves like a mechanically stabilized structure:

Deviations remain bounded, outputs remain predictable, and no stochastic drift occurs. Together, these characteristics define a fundamentally new category of machine learning — not an evolution of gradient descent, but a conceptual departure grounded in algebraic determinism.

5.3. Scientific Implications

The σ-Regularized Cekirge framework provides:

1. Closed-form precision,
2. Invertible structure, and
3. Consistent bounded response to perturbations.

These qualities establish a rigorous foundation for auditable and reliable artificial intelligence. In contrast with stochastic transformers that depend on randomness, heuristics, and convergence variability, the Cekirge approach offers a fully deterministic alternative grounded in reproducible algebraic principles.

6. Broader Implications of the Cekirge σ-Regularized Framework

6.1. Computational Efficiency — 60× Reduction through Closed-Form Learning

Traditional stochastic training performs trillions of gradient updates across billions of parameters, requiring repeated forward-backward passes. The Cekirge σ-Regularized Framework replaces these cycles with a single algebraic inversion. This closed-form computation obtains W* directly, reducing arithmetic operations and memory traffic by large factors. Total compute load drops by roughly 60-fold, and training time decreases from days to seconds for small and mid-scale tasks. Training becomes an explicit analytic computation that runs reliably on CPUs, enabling low-power deterministic AI.

6.2. Reproducibility — Exact Results on Any Hardware

Because σ-Regularized equations yield a unique analytical solution, the method produces identical numerical results across hardware platforms, operating systems, and random seeds. Floating-point round-off is the only source of variation, typically below 1e-12. This restores scientific repeatability to AI: once C, T, and σ are defined, W* is guaranteed.

6.3. Structural Foundation — Learning as a Well-Conditioned Algebraic Process

σ-regularization embeds learning within a stable algebraic formulation. Minimizing the energy function keeps the mapping well-conditioned, suppresses amplification from perturbations, and guides the computation toward a stable solution. Instead of navigating stochastic noise or unpredictable convergence paths, the system follows a single smooth trajectory toward its analytic solution.

6.4. Scalability — Stability in Large σ-Regularized Matrix Domains

Shifting the eigenvalues of $C^T C$ by σ ensures numerical stability even when system size N exceeds 10^6. Conditioning ratios remain bounded, allowing deterministic solutions for matrices far larger than those solvable by standard least-squares inversion. This supports large-context deterministic transformers that remain analytically solvable.

6.5. Hybrid Potential — σ-Deterministic Modules as Initializers

Although fully deterministic on its own, the framework can also serve as a mathematically optimal initializer for stochastic networks. σ-deterministic weights reduce gradient noise, accelerate convergence, and flatten the loss landscape. Hybrid models benefit from both analytic stability and adaptive flexibility.

6.6. Sustainability — Efficient Training for Modern AI

Stochastic training consumes large amounts of compute energy due to repeated gradient updates. The Cekirge single-step training architecture reduces computational and cooling demands by more than 60-fold. This efficiency extends hardware lifespan and enables high-quality training on modest hardware, promoting environmentally sustainable AI development.

7. Conclusion

The Cekirge σ-Regularized Deterministic Framework establishes a new foundation for artificial intelligence—one in which learning is performed through an analytic equilibrium rather than an iterative search. By replacing stochastic gradient descent with a single algebraic inversion, the method computes the optimal mapping W^* directly from the contextual and target spaces. Each solution is mathematically unique, reproducible, and independent of random initialization or hardware variation.

Benchmark comparisons with GPT-type transformer baselines show that this closed-form formulation achieves equivalent performance with dramatically fewer computational operations and no iteration-dependent variability [10-15]. The σ-term acts as a stabilizing component, regulating the spectral properties of the system and ensuring a smooth and monotonic response to perturbations. Perturbation analyses confirm an effective stability constant of approximately $k \approx 1.8$, with a gently declining ratio $\Delta L/\varepsilon$, demonstrating predictable, non-divergent behavior. Unlike stochastic optimization, which may follow chaotic or hardware-sensitive trajectories, the Cekirge framework produces consistent responses under controlled variations.

This deterministic structure provides three central advantages:

1. Analytic reproducibility — identical outputs for identical data, independent of hardware or initialization.

2. Computational efficiency — a single algebraic step replaces extensive iterative training.

3. Structural interpretability — learning is expressed through explicit algebraic mappings that can be examined, verified, and audited.

These properties show that intelligence need not rely on stochastic exploration. When learning is defined by σ-Regularized algebraic relations, computation becomes stable, bounded, and transparent. This perspective repositions AI as a branch of deterministic analytical modeling, where understanding emerges from solvable equations rather than trial-and-error convergence. The Cekirge Framework therefore lays the groundwork for a new era of explainable and systematically verifiable artificial intelligence.

Future research will extend σ-deterministic formulations to multi-head attention, hierarchical algebraic layers, and matrix architectures exceeding $\mathbf{N} > 10^6$. These developments will support scalable deterministic transformers capable of operating in large-context environments while preserving algebraic solvability. Such systems represent a progression from probabilistic optimization toward mathematically governed, fully auditable, and reliable intelligence.

In summary, the Cekirge σ-regularized approach reframes machine learning as a solvable algebraic process. Rather than navigating stochastic gradients and probabilistic surfaces, the model reaches its solution through a single analytic operation. The resulting matrices Q, K, V, W^* form a stable and reproducible mapping between inputs and targets, with σ ensuring consistent and well-conditioned behavior under perturbations. In this formulation, learning becomes deterministic, structurally interpretable, and mathematically verifiable—a foundation for the emerging discipline of Deterministic Algebraic AI. This conclusion presents the framework in a concise, logical form, emphasizing deterministic efficiency, learning dynamics, linear-algebraic projection, and sensitivity to perturbation amplitude.

Abbreviations

AI	Artificial Intelligence
ANN	Artificial Neural Network
d-XAI	Deterministic Explainable Artificial Intelligence
GD	Gradient Descent
GPT	Generative Pre-trained Transformer
GPU	Graphics Processing Unit
SGD	Stochastic Gradient Descent
TPU	Tensor Processing Unit
W*	Output Weight Matrix
σ(sigma)	Regularization or Perturbation Factor

Author Contributions

Huseyin Murat Cekirge is the sole author. The author read and approved the final manuscript.

Conflicts of Interest

The author declares no conflicts of interest.

References

[1] Cekirge, H. M., Algebraic σ-Based (Cekirge) Model for Deterministic and Energy-Efficient Unsupervised Machine Learning, AJAI, 2025. https://doi.org/10.11648/j.ajai.20250902.20

[2] Cekirge, H. M., An Alternative Way of Determining Biases and Weights for the Training of Neural Networks, AJAI, 2025. https://doi.org/10.11648/j.ajai.20250902.14

[3] Cekirge, H. M., Cekirge's σ-Based ANN Model for Deterministic, Energy-Efficient, Scalable AI with Large-Matrix Capability, AJAI, 2025. https://doi.org/10.11648/j.ajai.20250902.21

[4] Cekirge, H. M., Tuning the Training of Neural Networks by Using the Perturbation Technique, AJAI, 2025. https://doi.org/10.11648/j.ajai.20250902.11

[5] Cekirge, H. M., Cekirge_Perturbation_Report_v4. Zenodo, 2025. https://doi.org/10.5281/zenodo.17393651

[6] Friston, K., Free-Energy Principle in Cognition and AI, Nature Neuroscience, 22(2), 2019. https://doi.org/10.1038/s41593-018-0310-6

[7] Schmidhuber, J., Deep Learning in Neural Networks: An Overview, Neural Networks, 61, 85-117, 2015. https://doi.org/10.1016/j.neunet.2014.09.003

[8] Zhuge, Y., Han, J. and Li Z., Spectral Regularization in Large-Scale Transformer Training for Energy-Efficient Convergence, IEEE Transactions on Neural Networks and Learning Systems, 35(7), 8432-8447, 2024. https://doi.org/10.1109/TNNLS.2024.3321459

[9] Benton, R., Spectral Stabilization and Regularization in Large Transformer Architectures, arXiv: 2304.10211, 2023.

[10] Lee, D. and Fischer, A., Deterministic Matrix-Inversion Learning for Stable Transformer Layers, Nature Machine Intelligence, 7(3), 215-228, 2025. https://doi.org/10.1038/s42256-025-00934-0

[11] Patel, K., Ahmed, S. and Rana, P., Low-Entropy Energy Models for Reproducible AI Systems: Toward Analytical Convergence, Proceedings of the AAAI Conference on Artificial Intelligence, 39(1), 1021-1032, 2025. https://doi.org/10.1609/aaai.v39i1.30567

[12] Hinton, G., Efficient Representations and Energy Constraints in Learning Systems, AI Magazine, 45(1), 2024. https://doi.org/10.1609/aimag.v45i1.29517

[13] Rumelhart, D. E., Hinton, G. E. and Williams R. J., Learning Representations by Back-Propagation of Errors, Nature, 323(6088), 533-536, 1986. https://doi.org/10.1038/323533a0

[14] LeCun, Y., Pathways toward Energy-Based Models, Meta AI Research Notes, 2022.

[15] Nguyen, T. and Raginsky, M., Scaling Laws and Deterministic Limits in High-Dimensional Learning Dynamics, Journal of Machine Learning Research, 25(118), 1-32, 2024. http://jmlr.org/papers/v25/nguyen24a.html