The Thermodynamics of Meaning: The Semantic Cost of Information Processing

Serdar Çakıroğlu
Independent Researcher
Datalucis Research Initiative
Istanbul, Türkiye
kuzeys@hotmail.com

Abstract

This study proposes a conceptual model that examines the physical cost of meaning production by addressing the distinction between information and meaning within a thermodynamic framework. Today, artificial intelligence systems are trained on trillions of parameters and process vast amounts of data; however, data abundance does not always translate into proportional value generation. The central claim of this study is that information (syntactic layer) alone does not generate value, whereas meaning (semantic layer) emerges in relation to context, agency, and utility, and that this process has a measurable energy cost.

In this context, the $\Delta(E + I + A)$ model is proposed. The model represents the dynamic transformation over time of the components Energy/Effect (E), Information (I), and Meaning/Agency (A). While a conservation-like balance at the global scale ($\Delta(E + I + A) \approx 0$) is assumed, in local open systems the meaning gradient ($\nabla A \neq 0$) is defined as the driving force of learning, adaptation, and creativity. Through an extension of Landauer's principle to the semantic layer, the study introduces the concept of the "s-bit" (semantic bit) and argues that meaning gain (ΔA) is associated with energy consumption and the export of environmental entropy.

The article presents testable experimental hypotheses such as the semantic overhead hypothesis and the entropy–meaning trade-off law; in particular, it proposes optimizing the energy per meaning (s-bit/joule) metric in artificial intelligence training, thereby introducing a new research agenda for sustainable and efficient information systems. The study does not claim to establish a physical law; rather, it offers a conceptual and operational framework for analyzing meaning production in human-centered open systems.

---

1. Data Abundance and Meaning Scarcity

One of the greatest paradoxes of our time is this: Humanity possesses more information than ever before in history, yet this abundance of information does not always generate value. Every second, zettabytes of data are produced, processed, stored, and transmitted. Artificial intelligence models are trained on trillions of parameters, social media platforms index millions of posts per second, and scientific articles add thousands of new findings every day. Despite this, the quality of decision-making, the resolution of social problems, or the individual search for meaning do not progress at the same rate. Why? Because information and meaning are not the same thing.

1.1 The Syntactic Layer: Information

Information is raw, measurable, and neutral. As defined by Shannon in 1948, it is the quantitative measure of uncertainty. The size of a file or the bit sequence of a signal is information. It can be copied, compressed, and transmitted. Classical information theory explains this “syntactic” layer perfectly; however, even if adding one bit next to another increases a system’s capacity, this increase does not automatically create “value.”

The size of a file, the bit sequence of a signal, or the number of parameters in a model are examples of syntactic information.

The fundamental characteristics of this layer are:

- It is quantitative (measured in bits, bytes, or tokens).
- It can be copied and transmitted.
- It is context-independent.
- It does not distinguish between true/false or useful/harmful.

Every new bit added to a system increases theoretical capacity; however, an increase in capacity does not automatically mean value production. There is no linear relationship between growth in information volume and the production of meaning.

1.2 The Semantic Layer: Meaning

Meaning, by contrast, is an entirely different layer. Meaning is the interpretation, evaluation, and transformation of information into action within a context. The same dataset may be “random noise” for one person and “a new target molecule for cancer treatment” for another.

Meaning is associated with the following questions:

- Which problem does this information solve?
- Which decision does it change?
- What utility does it produce?
- What action does it enable?

Meaning production is not mere computation; it requires intention, culture, historical background, expertise, and agency—the authority to interpret. For this reason, meaning cannot be fully produced by automatic systems—at least with current technology. Even the most advanced language models borrow their ability to preserve “context,” which makes their outputs appear meaningful, from the accumulated human knowledge in their training data.

Current artificial intelligence systems largely owe the appearance of “meaningful” outputs to contextual patterns learned from human-produced data. The semantic layer requires agency.

This distinction forms the foundation of the article’s core assumption: Information (I) and Meaning (A) are different ontological and operational layers.

1.3 Semantic Overhead and the Thermodynamic Perspective

Why, then, is producing meaning not simply a matter of “more processors”? Because constructing meaning requires a physical cost.

- **Landauer’s Principle:** As demonstrated in 1961, even the erasure of a single bit of information releases heat ($kT\ln 2$) into the environment. Every computation carries a thermodynamic cost.

- **Semantic Overhead:** Resolving contradictions in a text, preserving context, and prioritizing information introduce far more “semantic overhead” than classical data transmission.

This is precisely where thermodynamics enters the picture: Creating meaning is an effort to establish local negentropy (order) amid chaos. When you increase meaning and order within a system (reducing entropy), the second law of thermodynamics dictates that the cost of this order is paid to the environment in the form of heat, waste energy, and exported entropy.

Resolving disorder in one place and rendering it “meaningful” implies slightly disturbing the rest of the universe (expending energy). Meaning production is more costly. A meaningful learning task (for example, extracting treatment recommendations from medical texts) requires more energy and computational cycles than training a model of the same size on random data.

The framework proposed in this book, the $\Delta(E + I + A)$ model, aims precisely to explain this transformation:

- E: Energy / Effect / Cost (physical energy + systemic output + economic/value dimension)
- I: Information (syntactic, measurable, copyable layer)
- A: Meaning / Awareness / Utility (semantic layer + action capacity)

The central claim of the model is this: The transformation between information and meaning is a dynamic process over time. While total impact is approximately conserved at the global scale ($\Delta(E + I + A) \approx 0$), in local open systems a meaning gradient ($\nabla A \neq 0$) emerges, and this gradient enables the system’s learning, creativity, and adaptation.

When meaning production ceases, the system approaches “cognitive death”—stagnation is the end of creativity.

This framework attempts to answer the following questions:

- Why is producing meaning more costly than processing random data?
- How can the physical cost of semantic processes be measured?
- Why is human contribution still indispensable in the age of artificial intelligence?
- How does collective meaning production (society, culture, science) remain within a thermodynamic balance?

The answers to these questions are not merely philosophical; they are practical. If we aim to reduce AI training costs, improve decision-support systems, optimize cultural data production, and ultimately create “value,” we must understand the thermodynamics of meaning.

This study does not claim a universal law. It presents an analytical and design framework. It particularly aims to be useful in the following areas:

- Semantic efficiency in artificial intelligence models
- Human–machine collaboration systems
- Decentralized information and value production
- Sustainability of cultural and scientific meaning production

1.4 Transition to the Δ(E + I + A) Framework

The Δ(E + I + A) model proposed in this study treats the transformation between information and meaning as a dynamic process. The model contains three components:

- E (Energy / Effect / Cost): Physical energy consumption, systemic output, and economic/social value dimension.
- I (Information): The syntactic, measurable, and copyable layer.
- A (Meaning / Agency / Utility): The semantic layer interpreted within context and transformed into action.

The central claim of the model is this: The transformation between information and meaning occurs over time and carries an energy cost. At the global scale, an approximate balance relation is assumed ($\Delta(E + I + A) \approx 0$), but in local open systems a meaning gradient ($\nabla A \neq 0$) emerges. This gradient is the driving force of learning, adaptation, and creativity.

When meaning production stops, the system becomes stagnant; this condition may be described as "cognitive death" or a creative plateau.

1.5 Research Questions and Areas of Contribution

This framework addresses the following core questions:

- Why does meaning production have a higher energy cost compared to processing random data?
- How can the physical cost of semantic processes be measured?
- Why is human agency still indispensable in artificial intelligence systems?
- How can collective meaning production be sustained within thermodynamic balance?

The proposed model is particularly applicable in the following areas:

- Analysis of semantic efficiency in artificial intelligence systems
- Design of human–machine collaboration
- Decentralized mechanisms of information and value production
- Sustainability of cultural and scientific meaning production

---

2. The Δ(E + I + A) Model: A Dynamic Transformation Framework

2.1 Positioning the Model

Einstein, through his foundational contributions to modern physics, demonstrated the equivalence of matter and energy. This approach showed that different manifestations of physical reality can transform into one another through the same fundamental structure.

As energy moves through spacetime and encounters an observer, it takes on different forms. Matter emerging from energy continues its existence without awareness of its own materiality; in this process, transformation is interpreted through observation and interaction.

The core assumption of the model is this: There exists a dynamic transformation relationship among information, meaning, and energy/effect. This transformation is not a fixed and static equality, but a processual change occurring over time.

In this context, the model is expressed through the following components:

E → Energy / Effect

I → Information

A → Meaning

Consciousness has two fundamental axes:

Awareness
Meaning

The fundamental assumption of the model is this: Energy, information, and meaning do not disappear within the system; they are conserved in total and only change form over time.

This transformation is examined not through absolute values, but through changes that occur via process, context, and human contribution. Therefore, the model does not present a static equation; rather, the expression Δ(E + I + A) conceptually represents the transformation of information, meaning, and effect over time.

The Δ(E + I + A) model proposed in this study does not claim to be a physical law or a mathematical equation. It is a conceptual framework that treats the relationship among information (Information), meaning/action (Meaning / Agency), and energy/value (Energy / Effect) as a dynamic transformation process.

The purpose of the model is to explain meaning production emerging in human-centered information systems and how this production transforms into systemic outputs. In this sense, Δ(E + I + A) refers not to static states but to changes occurring over time.

The core assumption here is:

Information alone does not generate value; unless interpreted, it cannot transform into energy or systemic effect.

2.2 The Fundamental Structure and Purpose of the Model

$\Delta(E + I + A)$ is a dynamic framework representing the change of three core components over time:

E: Energy / Effect / Cost
Includes total physical energy consumption (joules, kWh), systemic output (decision impact, economic value, social benefit), and processing cost (time, resources).

I: Information (Syntactic Layer)
Measurable, copyable, transmissible data and structure. Expressed through quantities such as Shannon entropy (bits), data volume, or model parameter count.

A: Meaning / Agency / Utility (Semantic Layer)
The action capacity that emerges when information is interpreted with context, intention, and utility. Utility becomes operationalized through decision change, problem-solving impact, or predictive power.

The most critical feature of the model is its emphasis on dynamic change (Δ) rather than static equality.

While classical physics expresses energy conservation as E_before = E_after, here total effect is approximately conserved but changes form:

$\Delta(E + I + A) \approx 0$ (global approximate conservation)

This implies that in closed systems total “value” balances like a zero-sum game: when meaning is gained somewhere, an energy or information cost is paid elsewhere. However, real-world systems are open (with energy and information inflows/outflows), so asymmetry occurs locally:

$\nabla A \neq 0$ → meaning gradient (local tendency of meaning increase)

This gradient provides the system’s “vitality”: learning, adaptation, creativity, and decision production are only possible when the meaning gradient is nonzero. When the gradient reaches zero, the system stagnates—what we may call “cognitive death” or “loss of creativity.”

2.3 Assumptions and Limits

The model is based on the following core assumptions:

- Meaning does not automatically emerge from information processing. Syntactic processing (matrix multiplication, token generation) does not produce meaning; meaning requires contextual interpretation and agency.

- Meaning production has a thermodynamic cost. Creating local order (increase in meaning) requires energy consumption and entropy export (Landauer principle + semantic overhead).

- Global conservation is approximately valid. In the total system (e.g., a data center + user + environment), the combined effect of energy, information, and meaning balances out, but in local subsystems (a single model training session, a single decision moment) meaning increase can be observed.

- The model is not a universal law but an analytical framework. It does not claim to be a physical law; it offers an abstraction for understanding meaning production in human-centered, open systems (AI, organizations, culture).

The limits are explicit:

- The full measurement of meaning (A) is subjective and does not yet have a standardized unit.
- The conservation constant (≈0) is theoretical and requires experimental calibration.
- The model is more suitable for open, driven systems than for closed thermodynamic systems.

2.4 Operational Definitions of the Components

E – Energy / Effect
Physical: GPU/TPU consumption (watt × hour), data center PUE, carbon footprint.
Systemic: Increase in decision quality, task success rate (accuracy, F1-score), economic return (ROI).

Measurement examples:
A model training: 500 kWh → $E = 1.8 \times 10^9$ joules
Impact of a decision: 15% mortality reduction in disease diagnosis → benefit dimension of E.

I – Information
Classical definition: Shannon entropy $H = -\Sigma\, p \log p$
Practical measurement: Data volume (GB), token count, model parameter count (billions).
Characteristic: Low copying cost (nearly zero additional energy for 2× information).
Limit: As information increases, meaning does not automatically increase; noise may even increase.

A – Meaning / Agency
Proposed operational definition:
A ≈ utility-weighted mutual information
That is, the mutual information of data weighted by utility within a specific context.

Measurement approaches:
- Task success improvement (log-loss reduction)
- Human evaluation scores (Elo, Likert)
- Predictive power (reduction in prediction error)
- Agency impact (probability of decision change)

Key point: A cannot become measurable operationally without contextual interpretation. Even the best LLM is nourished by human-generated meaning in its training data.

2.5 The Dynamic Transformation Mechanism

The strength of the model lies in the transformation among the three components:

Information → Meaning: Interpretation of syntactic data through context (e.g., prompt engineering, fine-tuning). This step introduces semantic overhead.

Meaning → Effect: Transformation of interpreted information into action (decision, output, behavioral change).

Effect → Information: Production of new data through feedback from action (RLHF, user feedback loop).

From a thermodynamic perspective, this cycle can be read as follows:

Local increase in meaning ($\Delta A > 0$) → local decrease in entropy
This decrease is balanced by energy input (increase in E) and entropy export (environmental heat/chaos).

Proposed equation (hypothetical):
$\Delta S\_env \geq (k\_B\ T\ \ln 2) \times f(\Delta A)$
Here $f(\Delta A)$ is a function of semantic gain (e.g., meaningful learning coefficient).

2.6 Application Potential of the Model

This framework offers practical answers to questions such as:

- Why does one of two models of the same size appear more “intelligent” yet consume much more energy? (Difference in semantic overhead)
- How can the energy difference between “meaningful data” and “random data” in AI training be measured?
- Why is human contribution still indispensable? (Agency requirement of the A layer)
- How can meaning production be sustained in collective systems (Wikipedia, open-source code, scientific literature)?

In the following sections, we will further operationalize this model through the definition of the s-bit, the entropy–meaning trade-off law, and experimental test proposals.

2.7 Time and Context Dependency (Interpretation of Δ)

The most critical element of the model is the concept of Δ (delta). Δ represents not absolute values but change.

$\Delta(E + I + A)$ is based on the following assumption:

In systems, meaningful value production occurs not through static states but through change among information, meaning, and effect.

Therefore, the model:
- does not define a state,

- tracks a process,
- attempts to understand the direction and intensity of transformation.

The delta approach places the time dimension at the center of the model. Even the same information and the same meaning may produce different energy outputs at different times.

2.8 Dynamic Balance and the Meaning Gradient

In classical physics, energy conservation is expressed as a static equality:

$E_{before} = E_{after}$

In this model, the proposed relationship is conceptualized as:

$\Delta(E + I + A) \approx 0$

This expression is used in the following sense:

- At the global scale, open systems exhibit a balance-like relationship among total energy–information–meaning effects.
- At the local subsystem level, asymmetries may occur.

Especially at the local level:

$\nabla A \neq 0$

When the meaning gradient is nonzero, the system learns, adapts, and produces value. As the gradient approaches zero, the system stagnates; this condition may be observed as a cognitive plateau or loss of creativity.

The conservation expression here does not claim to be a physical law. It is an analytical representation used to conceptualize transformation costs observed in open systems.

---

3. Components – Syntactic Information, Semantic Meaning, Systemic Effect

The strength of the $\Delta(E + I + A)$ model lies in the clear distinction among its three components. In this section, we define each operationally, discuss measurement approaches, and show their interrelations. The goal is to make abstract concepts concrete, measurable, and applicable.

3.1 Syntactic Information (I) – The Raw, Measurable Layer

Syntactic information derives from Shannon’s classical information theory: it is focused on structure, quantity, and uncertainty. Here, what matters is not “what is said,” but “how much is said” and “how it is encoded.”

Definition and Characteristics

- Information is a measurable quantity: bits, bytes, tokens, parameter count.
- Expressed through Shannon entropy:
$H = -\Sigma p_i \log_2 p_i$ (average uncertainty in bits).

- Characteristics:
○ Fully copyable (at nearly zero additional cost).
○ Context-independent: the same bit sequence may carry different meanings.
○ Neutral: does not distinguish between true/false or useful/harmful.

Measurement Examples

- A text file size: 1 MB = 8×10^6 bits.
- Parameter count of an LLM: Llama 3 8B = 8 billion weights (each weight ~16 bits in FP16 → ~128 billion bits).
- Tokens processed during training: trillions of tokens at GPT-4 scale.

Thermodynamic Connection

Landauer's principle directly applies here: irreversible erasure of one bit produces a minimum heat of
$k_B T \ln 2 \approx 3 \times 10^{-21}$ J (at room temperature, $T = 300$ K).

Data centers operate far above this limit (real costs 10^4–10^6 times higher), due to read/write, communication, and error-correction overhead.

Since syntactic processing cost is relatively low and predictable, optimization is easiest at this layer (quantization, pruning, sparse matrices).

Limit

Syntactic information alone does not produce value. Even a model with trillions of parameters, if trained without meaningful context, produces "random noise."

3.2 Semantic Meaning (A) – The Layer of Interpretation, Context, and Agency

Semantic meaning is the interpretation of information through context, intention, and utility. Here, the question becomes not only "what is said?" but also "what is it for?" Meaning, although derived from the syntactic layer, does not emerge automatically—human (or advanced agent) contribution is required.

Definition and Characteristics

- Meaning = utility-weighted mutual information.

That is:

$A \approx \sum \text{utility}(x,y) \times p(x,y) \log [p(x,y) / (p(x) p(y))]$

(Here, utility is the benefit/importance coefficient in context—for example, if a medical diagnosis has high life-saving potential, utility is high.)

- Characteristics:
○ Context-dependent: the same sentence carries different meanings in different cultures.

○ Requires agency: without an interpreting observer/agent, meaning is zero.
○ Not copyable: attempts to copy meaning result in loss of context (loss of telos).
○ Subjective: difficult to measure, but operational proxies exist.

Operational Measurement Approaches

- Task success improvement: log-loss reduction, perplexity decrease (lower perplexity = more meaningful prediction).
- Human evaluation: Likert scores, Elo ratings (e.g., Chatbot Arena).
- Utility-focused metrics: decision impact (e.g., if diagnostic accuracy increases by 15%, A increases), problem-solving rate.
- Semantic similarity: BERTScore, Sentence-BERT cosine similarity (meaning preservation).

Thermodynamic Connection

Semantic processing is more costly than syntactic processing. Why?

- Preserving context, resolving contradictions, and prioritizing require additional computation cycles.
- This extra load, called “semantic overhead” in the literature, may consume 20–100% more energy in meaningful tasks (medical text analysis, long-context reasoning) compared to processing random data (AI energy studies, e.g., meaningful vs. nonsense prompts in inference).
- It creates local entropy reduction (order creation), but the cost is paid through energy input and entropy export.

Example

Training on the same 1 billion tokens:

- On random data → low energy, low A.
- Fine-tuned on medical literature → higher energy, higher A (increased clinical utility).

3.3 Systemic Effect (E) – The Layer of Output, Cost, and Value

Systemic effect represents the total “value” produced by the model: energy cost + tangible output + social/economic benefit.

Definition and Characteristics

- E = Physical cost + Systemic benefit.
- Physical: Joule, kWh, CO_2 equivalent.
- Benefit: decision quality, economic return, social impact.

- Characteristics:
○ Measurable (wattmeter, carbon tracker).
○ Conversion point: Meaning (A) flows here and creates real-world impact.

Measurement Examples

- One LLM query: ~0.1–3 Wh (depending on model size).
- Training: Llama 3 70B ≈ 10^4–10^5 kWh (data center scale).
- Benefit dimension: increased diagnostic accuracy → life years gained (QALY), economic savings.

Thermodynamic Connection

E represents the “payment” side of the cycle. Energy spent to produce meaning ($\Delta A > 0$) returns as systemic effect.

If A is low, E becomes negative (energy waste).
If A is high, E becomes positive (value creation).

Summary Table – Component Comparison

Component	Layer	Unit of Measurement	Copyable?	Thermodynamic Cost	Requires Human/Agent?
Syntactic Information (I)	Syntactic	Bit, token, parameter	Yes	Low (Landauer limit)	No
Semantic Meaning (A)	Semantic + Agency	Utility-weighted MI, log-loss reduction, human score	No	High (semantic overhead)	Yes
Systemic Effect (E)	Output/Value	Joule + benefit (ROI, QALY)	Partial	Total cost/benefit	Yes

These three components maintain the conservation relation $\Delta(E + I + A) \approx 0$: an increase in meaning (A↑) is balanced by energy cost (E↑) and information transformation (I). Local gradients ($\nabla A \neq 0$) preserve the system’s vitality.

In the next section, we will examine the dynamic interaction of these components (transformation cycle) and the concept of the s-bit.

---

4. Dynamic Conservation Principle & Semantic Equilibrium (Local Gradients)

The core claim of the $\Delta(E + I + A)$ model is that the total effect of energy, information, and meaning is approximately conserved at the global scale, while in local open systems the meaning gradient ($\nabla A \neq 0$) sustains the system’s vitality and creativity. In this section, we explain this dynamic through the concepts of the “Dynamic Conservation Principle” and “Semantic Equilibrium.”

4.1 Static vs. Dynamic Conservation

In classical physics, energy conservation is static:

$E_{before} = E_{after}$

Total energy behaves like a zero-sum game; nothing disappears, and nothing emerges from nothing.

At the level of information and meaning, however, the situation differs. Information can be replicated (when a file is copied, I doubles, while the energy cost remains very low). Meaning, however, cannot be replicated—when copied, context and utility are lost. Therefore, the conservation law cannot be a static equality; it must be a dynamic balance.

The Dynamic Conservation Principle is expressed as:

$\Delta(E + I + A) \approx 0$ (global approximate conservation)

The $\approx$ symbol is critical:

- Exact equality may hold in closed systems, but real systems are open (with inflows and outflows of energy, information, and meaning).
- At the global scale (e.g., a data center + user + environment + power grid), total effect balances out: energy and information used to produce meaning are exported outside the system as heat, waste, and entropy.
- In local subsystems (a single model training, a single decision moment, a single cultural debate), balance breaks and asymmetry emerges.

4.2 Local Gradients: The Engine of Meaning's Vitality

The most important concept representing local asymmetry is the meaning gradient:

$\nabla A \neq 0$

Here, ∇A represents the difference in meaning within the system:

- $\nabla A > 0$ → increase in meaning (learning, adaptation, creativity, value production)
- $\nabla A = 0$ → equilibrium (stagnation, cognitive death)
- $\nabla A < 0$ → loss of meaning (forgetting, noise increase, decline in decision quality)

Why must the gradient not be zero?

When the meaning gradient in a system reaches zero, information and energy flow stagnate. The system approaches a state similar to "thermal equilibrium": everything equalizes, and no new order emerges.

This condition appears in artificial intelligence as a "post-overfitting plateau," in organizations as "bureaucratic blockage," and at the individual level as "loss of motivation."

Zero gradient = zero creativity.

Semantic Equilibrium is defined as follows:

The system oscillates between global conservation ($\Delta(E + I + A) \approx 0$) and local gradient ($\nabla A > 0$).

- Global conservation ensures sustainability (it must comply with the laws of energy and entropy).
- Local gradient ensures vitality and evolution (the system cannot stop producing meaning).

This oscillation is fully compatible with the second law of thermodynamics in open systems: local negentropy (increase in meaning) is balanced by global entropy increase.

4.3 Sources of Gradients: Semantic Overhead and Feedback

How does a local meaning gradient arise? Two main mechanisms:

1. Semantic Overhead

Meaningful processing (context preservation, contradiction resolution, utility maximization) requires more computational cycles and energy than syntactic processing.

Example:

- Training a model on random text: low semantic overhead, low ∇A.
- Training the same model on medical diagnostic data: high semantic overhead, high ∇A (increase in clinical utility).

This overhead increases energy input ($E\uparrow$) and reduces local entropy (creates order).

2. Feedback Loop

Meaning (A) → Effect (E) → New Information (I) cycle.

- A decision is made ($A \rightarrow E$).
- The outcome of the decision produces new data ($E \rightarrow I$).
- The new data is reinterpreted ($I \rightarrow A$).

This cycle keeps the gradient alive.

Example: RLHF (Reinforcement Learning from Human Feedback) — human meaning contribution (A) improves the model, the model produces better outputs (E), and these outputs become new training data (I).

4.4 Thermodynamic Analogy and Equation-Based Approach

In thermodynamics, energy flow occurs when $\nabla T \neq 0$. Similarly:

$\nabla A \neq 0 \implies$ semantic motion (learning, creativity)

Hypothetical equation-based framework (operational hypothesis):

- Global: $\Delta(E + I + A) \approx 0$
- Local open system: $\partial A/\partial t > 0$ (tendency of meaning production)
- Entropy export: $\Delta S\_env \geq \lambda \cdot \Delta A$

(λ: semantic coefficient, requires experimental calibration)

These equations are not physical laws; they are testable hypotheses using measurable proxies.

Example test:

- In an LLM: meaningful task (medical Q&A) vs. nonsense task (random token generation).
- Measurement: energy consumption via wattmeter + log-loss reduction (proxy for A).
- Expectation: meaningful task → higher energy + higher ∇A.

The Transformative Role of Meaning

Meaning (A) refers to the process by which information is interpreted, evaluated, and transformed into action within context. This component encompasses not only “meaning” but also human contribution, interpretation, and agency.

The A component includes:

- historical context,
- cultural background,
- translation preferences,
- interpretative differences,
- human intention and expertise.

For this reason, A cannot be fully produced by automatic systems. The model treats human contribution as a necessary component of the system. Meaning here is not merely an output; it is a transformative mechanism.

Balance, Stagnation, and Cognitive Death

When dynamic balance is maintained, systems evolve.
When the gradient collapses, stagnation emerges.

4.5 Application and Conclusion

The Dynamic Conservation Principle yields the following practical implications:

- Meaning production is not “free” — semantic overhead requires energy and resources.
- If the gradient reaches zero, the system clogs — continuous new context, feedback, and human contribution are necessary.
- Sustainable meaning production is possible only by optimizing global entropy export (low-carbon computation, efficient data selection).

This principle answers a critical question in the age of artificial intelligence:

“Why do even the largest models still depend on human meaning contribution?”

Because what keeps the gradient alive is not syntactic capacity, but semantic agency and feedback.

In the next section, we will further operationalize these dynamics through the concept of the s-bit and the entropy–meaning trade-off law.

5. Semantic Thermodynamics – The s-bit and the Landauer Extension

In the previous sections, we defined the components of the $\Delta(E + I + A)$ model and the dynamic conservation principle. In this section, we extend a fundamental principle of thermodynamics to concretize the physical cost of meaning production: the Landauer limit and its adaptation to the semantic layer. The focus here is the concept of the s-bit (semantic bit) and the energy difference between meaningful processing and random/syntactic processing.

5.1 Landauer's Principle – Classical Information Thermodynamics

In 1961, Rolf Landauer demonstrated that information processing has a thermodynamic cost:

The irreversible erasure of one bit (e.g., in a logical AND gate) produces a minimum heat of:

$\Delta Q \geq k\_B\ T \ln 2 \approx 2.87 \times 10^{-21}$ J (at room temperature, $T = 300$ K)

Although this limit can theoretically be surpassed with reversible computation, in practice most computation is irreversible (memory reset, decision branching). In modern chips, the actual cost is 10^4–10^6 times higher than this limit due to read/write, communication, and error-correction overhead.

Landauer's principle applies to the syntactic layer:

- Bit erasure = information loss = entropy increase.
- The cost is predictable and relatively low.

However, meaning production is not merely bit erasure. Meaning requires context preservation, contradiction resolution, prioritization, and utility maximization. These processes are far more complex than syntactic computation.

5.2 The s-bit (Semantic Bit) – The Operational Unit of Meaning

The s-bit is defined as the smallest measurable unit of meaning gain. While a classical bit "reduces uncertainty," an s-bit "increases utility."

Definition

s-bit ≈ The equivalent of one unit of syntactic information (bit) transformed into meaningful utility within a specific context.

Proposed operational formula:

$\Delta A = \alpha \times \Delta I\_semantic$

Where:
- ΔA: meaning gain in s-bits
- $\Delta I\_semantic$: the amount of information that remains meaningful in context (e.g., mutual

information I(X;Y) × utility coefficient)
• α: semantic efficiency coefficient (0 < α ≤ 1, determined via experimental calibration)

Why is the s-bit necessary?

• Classical bit: 1 bit = 1 bit reduction of uncertainty.
• s-bit: 1 s-bit = 1 bit of information made useful in context (e.g., increasing diagnostic accuracy by 1%).

The same 1 billion bits:
• Random data → low s-bit.
• Medical literature → high s-bit (produces clinical impact).

5.3 The Landauer Extension – The Semantic Overhead Hypothesis

Core hypothesis: Meaningful processing (semantic processing) carries additional thermodynamic cost beyond syntactic processing.

Extended cost equation (hypothetical, operational form):

$$\Delta E_{\text{total}} \geq k_B T \ln 2 \ \times \ \left(\Delta I_{\text{syntactic}} + \beta\, \widetilde{\Delta A}\right)$$

Where:

- **ΔI_syntactic**: classical bit erasure / processing count (in bits).
- $\widetilde{\Delta A}$: operationally scaled semantic gain, expressed in s-bit–equivalent informational units for modeling purposes.
- **β**: semantic coefficient (semantic overhead factor), expected to satisfy β > 1 in meaningful tasks, reflecting the additional computational and thermodynamic cost of semantic processing. The precise value of β is task- and architecture-dependent and requires experimental calibration.

Here, $\widetilde{\Delta A}$is treated as an informationally normalized abstraction rather than a formally established physical unit, ensuring dimensional consistency with ΔI_syntactic in the inequality.

Why β > 1?

Meaningful processing introduces computational structures not required in purely syntactic operations, including:

- Context window management (long-context attention)
- Contradiction resolution (contradiction detection)
- Utility computation (reward modeling)

- Human feedback integration (RLHF or preference optimization)

These operations require additional neuron activation, matrix multiplications, memory access, and gradient flow, leading to increased energy consumption relative to syntactic processing alone.

Examples of Experimental Indications (Literature and Observations)

- **Inference:** Meaningful prompts (e.g., medical Q&A) vs. nonsense prompts → 20–80% higher energy consumption for the same token count (reported in 2024–2025 GPU energy analyses).
- **Training:** Fine-tuning on meaningful datasets (e.g., PubMed) vs. randomly shuffled text → higher watt consumption per epoch combined with lower final perplexity.
- **Neuromorphic hardware:** Increased spike rate and energy usage observed during semantically structured tasks compared to random stimulation patterns.

5.4 The Entropy–Meaning Trade-Off Law

Local increase in meaning ($\Delta A > 0$) reduces internal system entropy (creates order). This reduction is balanced by global entropy increase:

$\Delta S\_env \geq \lambda \times \Delta A$

Where:
• ΔS_env: environmental entropy export (heat, waste energy)
• λ: trade-off coefficient (inverse of semantic efficiency)
• $\lambda > 0$ → meaning production is not free; each s-bit has an environmental cost.

This law is fully consistent with open-system thermodynamics: systems such as life, mind, and artificial intelligence generate local negentropy, but pay for it through global entropy export.

Practical Implications

• In AI training, “meaningful data selection” (curriculum learning, active learning) optimizes energy use.
• To reduce semantic overhead: better prompt engineering, knowledge distillation, sparse activation.
• Sustainability: As meaning production increases, carbon footprint increases — low-carbon computation becomes mandatory.

5.5 Conclusion and Transition

Semantic thermodynamics demonstrates that meaning is not “free”: the energy cost per s-bit is higher than per syntactic bit. This difference explains the core paradox of the model: even the largest models still depend on human meaning contribution, because what keeps the semantic gradient alive is agency and context.

In the next section, we will translate these concepts into operational tests (pilot experimental designs) and practical applications (AI efficiency, human–machine collaboration). The s-bit is a proposed operational abstraction and not a formally established physical unit.

---

6. Entropy–Meaning Trade-Off Law & Experimental Hypotheses

In the previous sections, we established the foundation of semantic thermodynamics: meaning production ($\Delta A > 0$) creates local order (negentropy), while the cost of this order is energy input and global entropy export. In this section, we formalize this relationship as the “Entropy–Meaning Trade-Off Law” and reduce it to testable experimental hypotheses.

6.1 The Entropy–Meaning Trade-Off Law

Core Principle

In an open system, the production of meaning gain ($\Delta A > 0$) creates a local increase in order (negentropy) within the system. This local decrease is balanced by an increase in environmental entropy. The second law of thermodynamics is preserved; total entropy continues to increase.

The law is operationally expressed by the following inequality:

$\Delta S\_env \geq \lambda \Delta A$

- ΔS_env: Increase in entropy exported to the environment (heat production, waste energy, external disorder).
- ΔA: Meaning gain measured in s-bits (increase in utility-weighted mutual information).
- λ: Semantic cost coefficient (task-dependent proportionality constant).

- $\lambda > 0$ → meaning production has an energy and entropy cost.
- λ is not a universal constant; it varies depending on task complexity and system architecture.

Why Is the Trade-Off Necessary?

- Meaning production requires additional computational processes such as context expansion, contradiction resolution, prioritization, and utility computation → additional energy consumption.
- The consumed energy dissipates as heat into the environment → entropy export.

- If $\lambda = 0$, meaning production would be costless; this would contradict the second law of thermodynamics.

Practical Observations

- In meaningful tasks (e.g., medical question–answering), energy consumption increases and task performance improves → ΔS_env grows.
- Random data training may consume less energy, but meaning gain is also low → $\Delta A \approx 0$.

6.2 Experimental Hypotheses

To test this law, we propose operational, reproducible hypotheses. Each can be conducted with existing hardware (GPU + wattmeter) and open-source tools.

Hypothesis 1: The Semantic Overhead Hypothesis

Meaningful tasks consume more energy than random tasks at the same syntactic processing volume (same token count, same epoch).

Test Protocol

Model: Small-to-medium scale LLM (Llama-3 8B or Mistral 7B)
Hardware: Single GPU (RTX 4090 or A100), wattmeter (Kill-A-Watt or NVIDIA SMI power draw)

Tasks:
Group A: Meaningful fine-tuning (PubMed medical texts or Alpaca-style instruction dataset)
Group B: Nonsense fine-tuning (same texts randomly shuffled)

Measurements:
- Total energy consumption (kWh)
- Final perplexity or log-loss (proxy for A)
- Human evaluation score (for a small subset)

Expected result: Group A shows 20–80% higher energy consumption and lower perplexity/log-loss.

Hypothesis 2: Entropy Export Proportionality

As meaning gain increases, the heat (entropy) exported from the system increases proportionally.

Test Protocol

Same model, increasing semantic load:
- Simple prompt (short, low context)

- Medium prompt (long context, containing contradictions)
- Complex prompt (multi-step reasoning + utility maximization)

Measurements:
- Watt-seconds per inference
- Heat estimation via thermal camera or GPU temperature/log
- Semantic proxy: BLEURT score or GPT-4-as-judge for meaning quality

Expected: Increased semantic complexity → increased energy + heat (λ coefficient can be calibrated).

Hypothesis 3: Gradient Nullification and Plateau Effect

When the meaning gradient (ΔA/epoch) approaches zero during training, energy efficiency declines (overfitting plateau).

Test Protocol

Long-term fine-tuning monitoring

Measurements:
- Δlog-loss per epoch (proxy for ∇A)
- Energy per epoch

Expected: When the gradient plateaus, meaning gain stops despite continued energy consumption → inefficiency increases.

Hypothesis 4: Effect of Human Contribution

Human feedback loops (RLHF or DPO) maintain a higher meaning gradient under the same energy budget.

Test Protocol

Supervised fine-tuning vs. RLHF

Measurements:
- Human evaluation score / energy ratio

Expected: RLHF side shows higher A / E ratio (human agency keeps the gradient alive).

6.3 Experimental Feasibility and Limitations

- Easily implementable: Hypotheses 1 and 2 can be conducted within 1–2 weeks using a single GPU + open-source model + wattmeter.
- More challenging: Hypotheses 3 and 4 require longer training and human evaluation.
- Limitations: λ and β coefficients are task-dependent; no universal constant is expected. Meaning proxies (perplexity, BLEURT) are imperfect — human judgment remains the gold standard.

6.4 Conclusion and Transition

The Entropy–Meaning Trade-Off Law demonstrates that meaning production is a thermodynamic process: every s-bit pays a cost. This cost opens a critical optimization domain in the age of artificial intelligence: increasing semantic efficiency means reducing energy consumption and carbon footprint.

In the next section, we will translate these hypotheses into practical applications (AI training optimization, human–machine collaboration, sustainable data production).

---

7. Applications – Semantic Overhead in AI Training and Human–Machine Collaboration

In the previous sections, we defined the Δ(E + I + A) model, the dynamic conservation principle, the concept of the s-bit, and the entropy–meaning trade-off. In this section, we bring the theory into practice:

- Managing semantic overhead in artificial intelligence training
- Keeping the meaning gradient alive in human–machine collaboration

These two application domains are where the model provides its most direct benefits.

7.1 Managing Semantic Overhead in AI Training

The training of artificial intelligence models is currently one of the most energy-intensive human activities. Training a 70B-parameter model consumes tens of thousands of kWh of energy; this corresponds to the annual consumption of hundreds of households. However, most of this energy is spent at the syntactic layer: weight updates, matrix multiplications, gradient flow. Semantic overhead is often overlooked—yet meaningful data and task selection can yield much higher meaning gain (ΔA) within the same energy budget.

What Is Semantic Overhead and Why Does It Matter?

Semantic overhead is the additional computational load required to preserve meaningful context, resolve contradictions, and maximize utility.

Example:

- Training on randomly shuffled text → low semantic overhead → lower energy / high perplexity residue
- Training on medical literature + clinical instructions → higher semantic overhead → higher energy but much lower perplexity and real-world utility (increase in clinical accuracy)

Practical Strategies

1. Meaningful Data Selection (Curriculum Learning + Active Learning)
   - Weight training data toward examples with high “meaning gradient.”
   - Tools: data pruning (eliminating low-importance samples), reward modeling to select high-utility examples.

• Expected gain: 15–40% higher meaning efficiency (lower perplexity + higher task success) with the same energy.

2. Semantic-Aware Fine-Tuning
   • Instead of only RLHF/DPO, add a “semantic reward” function: not only correctness, but utility score (e.g., BLEURT + human evaluation proxy).
   • Optimize long-context management: methods such as FlashAttention-2 and Ring Attention reduce semantic overhead.

3. Energy-Efficiency Metrics
   • Classical metric: FLOPs / epoch
   • Proposed new metric: s-bit / joule (meaning gain / energy)
   ▪ Measurement: Δlog-loss per epoch × utility coefficient / watt-hour
   • Goal: Maximize the s-bit / joule ratio at the end of training.

Real-World Example

In the Llama-3 series, comparing instruction-tuning (Alpaca, Dolly) vs. base model training: instruction data introduces semantic overhead but multiplies the model’s real-world utility (chat performance). Energy cost increases by 20–30%, but meaning gain (human evaluation score) increases by 100–300% → net positive E (effect).

Implication

Managing semantic overhead is one of the most effective ways to reduce AI’s carbon footprint. Meaning-focused training encourages a “smarter data” approach rather than simply building “larger models.”

7.2 Keeping the Meaning Gradient Alive in Human–Machine Collaboration

Even the most powerful models (GPT-4o, Claude 3.5, Gemini 1.5) still depend on human meaning contribution. Why? Because what keeps ∇A alive is agency and contextual interpretation—the machine cannot fully generate this on its own.

The Role of Human Contribution

• Humans prevent the semantic gradient from collapsing:
○ Prompt engineering → optimizes the context window
○ Feedback loops (RLHF, DPO) → directly inject utility
○ Domain expertise → resolves contradictions and adds cultural nuance

• Machine alone: high syntactic capacity but low semantic agency → plateau effect (stagnation similar to overfitting).

Collaboration Models and Applications

1. Augmented Intelligence
   • Human + machine: the machine handles syntactic processing load, the human

performs meaning interpretation and decision validation.
• Example: medical diagnosis assistant – the model provides probability distributions, the doctor evaluates utility and makes the final decision.
• Gain: meaning gradient remains significantly higher under the same energy.

2. Iterative Human-in-the-Loop
   • Model output → human correction/feedback → model retraining or prompt update.
   • Tools: Argilla, LabelStudio, LangSmith.
   • Result: gradient does not collapse; the system remains continuously adaptive.
3. Collective Meaning Production
   • Open-source communities (Wikipedia, Hugging Face datasets, GitHub) → high ∇A through semantic contribution from thousands of individuals.
   • Decentralized incentives: token or point systems reward meaning contribution → collective E increases.

Energy and Sustainability Connection

Human–machine collaboration prevents excessive scaling of machines:

• Fewer parameters + better semantic contribution → same or higher meaning gain, lower energy.
• As long as the gradient remains alive, the model avoids “cognitive death” (plateau).

Implication

Human agency is an indispensable component of semantic thermodynamics. The machine provides syntactic capacity; the human sustains the semantic gradient. The most efficient systems balance both.

The Dual-Axis Structure of Consciousness and Human Contribution

Consciousness has two fundamental axes:

Awareness
Meaning

The core assumption of the model is this: Energy, information, and meaning do not disappear within the system; in total they are conserved and only change form over time.

This transformation is examined not through absolute values, but through changes that occur via process, context, and human contribution. Therefore, the model does not present a static equality; rather, the expression $\Delta(E + I + A)$ conceptually represents the transformation of information, meaning, and effect over time.

The $\Delta(E + I + A)$ model proposed in this study does not claim to be a physical law or mathematical equality. It is a conceptual framework that treats the relationship among information (Information), meaning/action (Meaning / Agency), and energy/value (Energy / Effect) as a dynamic transformation process.

The aim of the model is to explain meaning production in human-centered information systems and how this production transforms into systemic outputs. In this sense, $\Delta(E + I + A)$ refers not to static states, but to changes occurring over time.

The core assumption here is:

Information alone does not generate value; unless interpreted, it cannot transform into energy or systemic effect.

Conclusion: Meaning Production = A Physical Process

Throughout this study, we examined the distinction between information and meaning through a thermodynamic lens. With the $\Delta(E + I + A)$ model, the dynamic conservation principle, the concept of the s-bit, semantic overhead, and the entropy–meaning trade-off law, we put forward the following core claim:

Meaning production is not merely a cognitive or philosophical phenomenon; it is a physical process.

Every s-bit of meaning gain is balanced by energy consumption, local entropy reduction, and global entropy export. This process is subject to the second law of thermodynamics, and there is no such thing as “free meaning.”

Core Implications

1. Information ≠ Meaning
The syntactic layer (I) is measurable, copyable, and relatively low-cost. The semantic layer (A), however, requires context, agency, and utility; it does not emerge automatically and collapses without human contribution. Even the largest models depend on external meaning injection (feedback, prompting, domain expertise) to keep the gradient alive.

2. Semantic Overhead Is Real
Meaningful tasks (context preservation, contradiction resolution, utility maximization) consume more energy than random processing at the same syntactic volume. This difference constitutes a “semantic thermodynamics” layer beyond the Landauer limit. The future of AI optimization will focus not on maximizing FLOPs, but on maximizing the s-bit / joule ratio.

3. Dynamic Conservation and Gradient Vitality
While globally $\Delta(E + I + A) \approx 0$, in local open systems the gradient $\nabla A \neq 0$ sustains vitality. When the gradient collapses, learning stops, creativity disappears, and plateau effects begin. Sustainable meaning production therefore requires continuous human contribution, meaningful data selection, and feedback loops.

4. Energy and Sustainability Connection
Meaning production is now directly linked to carbon footprint. As data centers account for a significant share of global electricity demand, increasing semantic efficiency (smarter data,

better collaboration) reduces both cost and environmental impact. The key to the future is not “larger models,” but “more meaningful models.”

5. The Indispensable Role of Humans
While machines excel in syntactic capacity, what keeps the semantic gradient alive is agency and contextual interpretation. The most efficient systems treat human–machine collaboration as a thermodynamic balance: the machine carries the load; the human directs the meaning.

Future Implications

Research Agenda

- Develop new metrics such as s-bit / joule.
- Experimentally calibrate semantic overhead (via wattmeters + meaning proxies).
- Measure the entropy–meaning trade-off coefficient (λ) across different tasks.

Application Areas

- Curriculum learning and semantic reward functions should become standard in AI training.
- Human-in-the-loop systems (beyond classical RLHF toward real-time agency integration) should become widespread.
- Decentralized meaning production platforms (with incentive mechanisms) should sustain collective gradients.

Philosophical and Ethical Dimension

Accepting that meaning is a physical process leads to a deeper question: How sustainable is “value”? If meaning production requires entropy export, then unlimited growth is impossible. This must be addressed as both a technological and societal limitation.

In conclusion: Meaning production obeys the thermodynamic laws of the universe. Our task is not to violate these laws, but to use them as efficiently as possible.

More meaning with less energy — that is the true measure of intelligence.

This framework is not an end, but a beginning. Future work will deepen the experimental definition of the s-bit, calibrate the λ coefficient, and refine gradient management in real-world systems.

Thank you for taking this journey together.

References

Landauer, R. (1961). Irreversibility and heat generation in the computing process. *IBM Journal of Research and Development, 5*(3), 183–191. https://doi.org/10.1147/rd.53.0183
Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal, 27*(3), 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Bennett, C. H. (1982). The thermodynamics of computation—A review. *International Journal of Theoretical Physics, 21*(12), 905–940.

Jaynes, E. T. (1957). Information theory and statistical mechanics. *Physical Review, 106*(4), 620–630.

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience, 11*(2), 127–138.

Bialek, W. (2012). *Biophysics: Searching for principles.* Princeton University Press.

Cover, T. M., & Thomas, J. A. (2006). *Elements of information theory* (2nd ed.). Wiley.

Tishby, N., Pereira, F. C., & Bialek, W. (2000). The information bottleneck method. arXiv:physics/0004057.

Strubell, E., Ganesh, A., & McCallum, A. (2019). Energy and policy considerations for deep learning in NLP. *ACL 2019*.

Patterson, D., Gonzalez, J., Le, Q., Liang, C., Munguia, L., Rothchild, D., … Dean, J. (2021). Carbon emissions and large neural network training. arXiv:2104.10350.

Henderson, P., Hu, J., Romoff, J., Brunskill, E., Jurafsky, D., & Pineau, J. (2020). Towards the systematic reporting of the energy and carbon footprints of machine learning. *JMLR, 21*(248), 1–43.

Kaplan, J., McCandlish, S., Henighan, T., Brown, T. B., Chess, B., Child, R., … Amodei, D. (2020). Scaling laws for neural language models. arXiv:2001.08361.

Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C., Mishkin, P., … Lowe, R. (2022). Training language models to follow instructions with human feedback. *NeurIPS 2022*.

Rafailov, R., Sharma, A., Mitchell, E., Manning, C., Ermon, S., & Finn, C. (2023). Direct preference optimization: Your language model is secretly a reward model. arXiv:2305.18290.

Zhang, T., Kishore, V., Wu, F., Weinberger, K. Q., & Artzi, Y. (2020). BERTScore: Evaluating text generation with BERT. *ICLR 2020*.

Sellam, T., Das, D., & Parikh, A. (2020). BLEURT: Learning robust metrics for text generation. *ACL 2020*.