# Semantic Dynamics: Studying the Thermodynamics of Semantic Particles

Omar Cusma Fait

July 2025

**Abstract**

We introduce *Semantic Dynamics*, a framework that interprets the evolution of token sequences in large language models (LLMs) as a thermodynamic system. Semantic embeddings define an effective energy landscape, and token generation follows stochastic trajectories in a latent state space. Using tools from classical mechanics and statistical thermodynamics, we derive interpretable quantities, such as temperature, potential energy, and kinetic energy, from model behavior. This enables the diagnosis and mitigation of pathological dynamics such as looping or stuck states. Our approach offers a practical diagnostic tool for estimating when a model is trapped in a semantic potential well and how to escape it by fine-tuning an effective temperature, directly addressing repetition and periodic token cycles in LLMs.

By modeling linguistic evolution as the motion of a "semantic particle" through a continuous embedding space, we unlock a physical interpretation of meaning change, where fluctuations in semantics resemble physical processes such as diffusion, inertia, and thermal agitation. This analogy allows us to quantify coherence, diversity, and stability in generated text using thermodynamic observables.

## 1 Introduction

Consider a corpus of text represented as a sequence of tokens $\boldsymbol{v} = (v_t)$, where each $v_t \in \mathcal{T}$ is drawn from a discrete vocabulary. Let $f : \mathcal{T}^N \to \mathcal{E}$ be a semantic embedding function that maps a window of $N$ consecutive tokens $\boldsymbol{v}[t : t + N]$ to a point $q_t \in \mathcal{E} \cong \mathbb{R}^d$ in a continuous semantic embedding space $\mathcal{E}$. This point $q_t$ captures the meaning of the local context carried by the sliding window at time $t$. From now on, the dependency on $\boldsymbol{v}$ is omitted but always implied.

The central insight of this work is that the discrete sequence of semantic embeddings $\boldsymbol{q} = (q_t)$ can be treated not just as isolated points but as partial observations of an underlying continuous trajectory, similar to tracking the position of a particle moving through an abstract space of meaning. This opens the door to analyzing linguistic dynamics using the full machinery of classical mechanics and statistical thermodynamics.

### 1.1 Key Idea #1: Extending the Semantic Trajectory to the Continuum

We treat the discrete sequence of embeddings $\boldsymbol{q} = (q_t)$ as partial observations of an underlying continuous trajectory $q(t) : \mathbb{R}^+ \to \mathcal{E}$. We can imagine to interpolate the discrete embeddings

$q_t = f(\boldsymbol{v}[t : t+N])$ obtained via a sliding window, to get a smooth curve $q(t)$. (see the *Continuum Semantic Trajectory Hypothesis*).

Differentiating this trajectory yields the *semantic velocity*:

$$\dot{q}(t) = \frac{d}{dt} q(t) \in T_q\mathcal{E},$$

which captures the instantaneous rate and direction of semantic change—the "flow of meaning" at position $t$. The velocity leads to the *momentum*:

$$p(t) = \frac{\partial L}{\partial \dot{q}}(q(t), \dot{q}(t), t) \in T_q^*\mathcal{E},$$

where the Lagrangian $L$ implicitly depends on a notion of *semantic inertia* $m$. Although $m$ may vary—scientific texts may resist change more than poetic ones—in the rest of the article, the quantities that depend linearly on mass are to be understood *per unit mass*, and *mass* is therefore omitted (see *Constant Inertia Hypothesis*).

This lifting of discrete tokens into a continuous dynamical system allows us to compute momentum and unlock the full suite of operators from Lagrangian mechanics. As we shall see, this geometric perspective transforms abstract language into a physical-like process governed by energy, force, and entropy. It is also important to notice that we are not required to actually perform the interpolation at any point, as it is only an abstraction that connects the *semantic embeddings* to *Statistical mechanics*.

## 2 Lagrangian Picture: Velocity and the Tangent Bundle

In the Lagrangian formulation, the state at time $t$ is given by:

$$(q(t), \dot{q}(t)) \in T\mathcal{E},$$

where $T\mathcal{E} = \bigsqcup_{q \in \mathcal{E}} T_q\mathcal{E}$ is the *tangent bundle* of the embedding space. This is the natural space for velocity-based dynamics, which we call the *semantic state space*.

### 2.1 Key Idea #2: The Semantic Particle

The trajectory $t \mapsto (q(t), \dot{q}(t))$ describes a "semantic particle" moving through $T\mathcal{E}$, analogous to a physical particle in a potential landscape shaped by semantics. This picture closely resembles the idea of a particle evolving under forces derived from meaning coherence and contextual stability.

This analogy opens the door to analyzing linguistic dynamics using tools from statistical mechanics—energy, entropy, temperature, and diffusion—as physical-like processes. Fluctuations in meaning, topic shifts, and even stylistic variation can be interpreted as manifestations of kinetic energy, thermal agitation, and drift in a high-dimensional space of ideas.

## 3 Hamiltonian Picture: Momentum and the Cotangent Bundle

Equip $\mathcal{E}$ with a Riemannian metric $g$, typically the Euclidean inner product $g_q(u, v) = u^\top v$. The momentum is then the covector:

$$p(t) = g_{q(t)}(\dot{q}(t), \cdot) \in T^*_{q(t)}\mathcal{E},$$

which identifies $p(t)$ with a linear functional on tangent vectors. The full state now lives in the *cotangent bundle*:

$$(q(t), p(t)) \in T^*\mathcal{E},$$

called the *semantic phase space*—the domain for Hamiltonian dynamics.

Although $p(t)$ and $m\,\dot{q}(t)$ coincide numerically under the Euclidean metric, they live in dual spaces: velocity in $T\mathcal{E}$, momentum in $T^*\mathcal{E}$. This distinction is crucial when $\mathcal{E}$ is curved or equipped with a non-trivial metric (e.g., Fisher information), which encodes the sensitivity of meaning to context perturbations.

If we replace the Euclidean metric with the Fisher information metric derived from the underlying language model, then $g_q$ encodes how sensitive the meaning is to small changes in context. In this case, momentum becomes curvature-aware, yielding a more faithful representation of semantic dynamics in non-uniform embedding spaces.

## 4 Recovering the Thermodynamic Quantities

With the geometric structure in place, we define the physical analogs using the canonical ensemble. Starting from the density $\rho(q)$, we derive all thermodynamic quantities—temperature, energy, entropy, pressure—as interpretable measures of linguistic behavior.

### 4.1 Key Idea #3: Connection to the Canonical Ensemble

We postulate that the system is in thermal equilibrium, allowing us to define a canonical ensemble over semantic states. The probability density $\rho(q)$ plays a *foundational* role: it quantifies how frequently different regions of $\mathcal{E}$ are occupied. High-density regions correspond to common, coherent, or stylistically typical meanings (e.g., standard syntactic patterns, frequent topics), while low-density areas represent rare, idiosyncratic, or disfluent constructions.

Thus, $\rho(q)$ serves as a direct proxy for *semantic plausibility*, and through the relation $V(q) = -\frac{1}{\beta}\log\rho(q)$, it defines the underlying potential landscape that governs the motion of the semantic particle.

#### *4.1.1 Semantic Density $\rho(q)$*

The probability density $\rho(q)$ is empirically estimated from the sequence of sliding window embeddings $q_t = f(\boldsymbol{v}[t : t+N])$, treated as samples from an unknown distribution. This density reflects the empirical likelihood of encountering a particular semantic state $q$ throughout the text.

High-density regions are "semantic basins"—stable, meaningful configurations , close to the topic of the text. On the other hand, most of the space is a low-density region, suggesting the idea of a *localized cloud of embedding vectors*.

To empirically estimate $\rho(q)$:

1. Compute embeddings: $q_t = f(\boldsymbol{v}[t : t+N])$.

2. (Optional) Apply dimensionality reduction (e.g., PCA, UMAP, autoencoders).
3. Estimate $\rho(q)$ using KDE, GMM, or $k$-NN density estimation.

When applying this framework, it is imperative to verify that the metric and density choices do not arbitrarily change the physics of the system. Likelihood-based analysis can help validate the estimated distribution under the assumption of equilibrium.

### 4.2 Key Idea #4: Effectively Estimating $\rho$

To obtain a robust estimate of $\rho(q)$ by adopting a *hybrid modeling approach*. We decompose the embedding space into two subspaces: the top $k$ principal components—capturing the dominant semantic directions—and the remaining orthogonal components, which often represent noise or less structured variation. The density is then modeled as a product:

$$\rho(q) \approx \rho_{\text{dom}}(q_{||}) \cdot \rho_{\text{noise}}(q_{\perp}),$$

where: - $q_{||} = \pi_k(q)$ denotes the projection onto the first $k$ components, and $\rho_{\text{dom}}$ is a flexible, non-parametric density estimate (e.g., KDE or GMM) over this low-dimensional subspace, - $q_{\perp}$ is the residual component, assumed to be approximately Gaussian:

$$\rho_{\text{noise}}(q_{\perp}) = \mathcal{N}(q_{\perp} \mid \mu, \sigma^2 I).$$

This factorization leverages accurate modeling where it matters most, while avoiding overfitting in high-dimensional tails, where the effects of Brownian motion are dominant. It enables meaningful density estimation even when full-dimensional resolution is impractical, ensuring that thermodynamic quantities remain well-grounded in the intrinsic structure of the semantic dynamics.

### 4.3 Dimensionality $d$

The ambient space is $\mathbb{R}^d$, but, noise aside, the dynamics should generally be confined to a lower-dimensional manifold $\mathcal{M} \subset \mathbb{R}^{d_{\text{eff}}}$ due to linguistic constraints—grammar, topic coherence, and style.

Define a smooth map $\pi : \mathcal{E} \to \mathcal{M}$. The projected trajectory $q_{\mathcal{M}}(t) = \pi(q(t))$ evolves in $\mathcal{M}$. The effective dimension $d_{\text{eff}} = \dim(\mathcal{M})$ replaces $d$ in the thermodynamic formulas, mitigating the curse of dimensionality. The ideal cut-off point for $d_{\text{eff}}$ may be estimated through *PCA*.

Using $d_{\text{eff}}$ guarantees that thermodynamic quantities reflect only the dynamically active degrees of freedom—the true number of independent ways in which meaning can evolve. Throughout, $d$ should be interpreted as $d_{\text{eff}}$, regardless of the nominal size of the embedding space. It is to be expected, though, that the *Brownian behavior* of the *semantic particle* is non-negligible even in the low-variance directions, in which case the correct approach is to scrap dimensionality reduction altogether.

## 4.4 Semantic Volume $\mathcal{V}_{\text{sem}}$

$\mathcal{V}_{\text{sem}}$ represents the effective extent of $\mathcal{E}$ explored by the semantic particle:

$$\mathcal{V}_{\text{sem}} = \int_{\mathcal{E}} dq\, \chi_\epsilon(q),$$

where $\chi_\epsilon(q) = 1$ if $\rho(q) > \epsilon$, else 0. Alternatively, $\mathcal{V}_{\text{sem}}$ can be the volume of $\{q \mid V(q) \leq E\}$ for energy $E$.

Unlike physical volume, $\mathcal{V}_{\text{sem}}$ generalizes the notion of "available space" to the $d$-dimensional abstract space of meaning. A large $\mathcal{V}_{\text{sem}}$ indicates broad exploration—diverse topics or styles; a small $\mathcal{V}_{\text{sem}}$ suggests focused discourse.

This generalization is mathematically consistent with statistical mechanics, where phase-space volumes are routinely defined in high-dimensional spaces. The semantic volume plays the same thermodynamic role as the physical volume: it serves as the conjugate variable of pressure.

### *4.4.1 Algorithm for $\mathcal{V}_{sem}$*

1. Choose threshold $\epsilon > 0$.
2. Estimate region where $\rho(q) > \epsilon$ (via parametrization or Monte Carlo).
3. Compute:

$$\boxed{\mathcal{V}_{\text{sem}} = \int_{\mathcal{E}} dq\, \chi_\epsilon(q)}$$

## 4.5 Potential Energy $V(q)$

The potential $V(q)$ represents a semantic landscape that guides the dynamics of meaning in text. It is a scalar field in $\mathcal{E}$ representing "semantic landscape" features—e.g., topic attractors, conceptual basins, or stylistic preferences. It can be recovered using the empirical density of the discrete semantic vectors in the corpus (e.g., via *kernel density estimation*). Under the Gibbs-Boltzmann hypothesis:

$$V(q, \beta) = -\frac{1}{\beta} \log \rho(q),$$

with $\beta = 1/T$, and $k_B = 1$ (natural units). High-density regions appear as low-potential "semantic basins," while rare constructions sit in high-potential hills.

### *4.5.1 Recipe for $V(q)$*

1. Compute $q(t)$ via sliding window and interpolation.
2. Estimate $\rho(q)$ (e.g., with *KDE*).
3. Compute:

$$\boxed{V(q, T) = -T \log \rho(q)}$$

### 4.6 Hamiltonian

Kinetic energy (velocity form):

$$E_{\text{kin}}(t) = \frac{1}{2} m |\dot{q}(t)|_g^2.$$

Momentum form:

$$E_{\text{kin}}(t) = \frac{1}{2m} |p(t)|_{g^{-1}}^2.$$

Thus, the Hamiltonian is:

$$H(q,p) = \frac{1}{2m} |p|_{g^{-1}}^2 + V(q).$$

### 4.7 Partition Function $Z(\beta)$

To transition from deterministic dynamics to statistical behavior, we introduce the *partition function*, the cornerstone of equilibrium statistical mechanics. It aggregates the contributions of all possible semantic states $(q,p)$, weighted by their likelihood under the Hamiltonian $H(q,p)$.

In thermal equilibrium:

$$Z(\beta) = \int_{T^*\mathcal{E}} dq\, dp\, e^{-\beta H(q,p)}.$$

With Euclidean metric and factorized Hamiltonian:

$$Z(\beta) = \left( \int dp\, e^{-\beta |p|^2/(2m)} \right) \left( \int dq\, e^{-\beta V(q)} \right).$$

The momentum integral is Gaussian:

$$\int dp\, e^{-\beta |p|^2/(2m)} = \left( \frac{2\pi m}{\beta} \right)^{d/2}.$$

So:

$$Z(\beta) = \left( \frac{2\pi m}{\beta} \right)^{d/2} \int_{\mathcal{E}} dq\, e^{-\beta V(q)}.$$

Using the thermal de Broglie wavelength $\lambda_B = \sqrt{\beta/(2\pi m)}$:

$$Z(\beta) = \frac{1}{\lambda_B^d} \int_{\mathcal{E}} dq\, e^{-\beta V(q)}.$$

### 4.8 Free Energy $F(\beta)$

From the partition function, we derive the *Helmholtz free energy* of the *gas of semantic embeddings*, which governs the thermodynamic balance between energy and uncertainty:

$$F(\beta) = -\frac{1}{\beta} \log Z(\beta) = \langle H \rangle - TS.$$

It balances semantic coherence (low $V$) and diversity (high $S$)—a natural trade-off between staying on topic and exploring related ideas.

### 4.8.1 Recipe for $F(T)$

1. Estimate $Z(T)$ (via parametrization or Monte Carlo).
2. Compute:

$$\boxed{F(T) = -T \log Z}$$

## 4.9 Average Kinetic Energy $\langle K \rangle$

To measure the average kinetic energy $\langle K \rangle$, we usually calculate the instantaneous kinetic energy at each point along the semantic trajectory and then average these values. However, the embedding vector is subject to *Brownian motion*. We postulate that the semantic particle follows this *SDE* in every dimension:

$$\boxed{\begin{aligned} dx_t &= v_t\,dt, \\ dv_t &= -\gamma\, v_t\,dt \; - \; k\,x_t\,dt \; + \; \sqrt{2\,\gamma\,T}\;dW_t. \end{aligned}}$$

In this equation, $\gamma$ is a positive coefficient of friction, and can be estimated directly from the embeddings $q_t$. The other parameter, $k$, is the coefficient of resonance of the quadratic potential in that dimension, and it can be estimated directly from $V(q)$.

### 4.9.1 Recipe for $\gamma$

1. Prepare the centered velocities:

$$\dot{q}_t = q_{t+1} - q_t$$
$$v_t = \dot{q}_t - \langle \dot{q} \rangle$$

2. Compute the lag-1 autocorrelation:

$$\rho_1 \; = \; \frac{\displaystyle\sum_{n=0}^{N-2} v_n\, v_{n+1}}{\displaystyle\sum_{n=0}^{N-1} v_n^2}$$

3. Compute to a damping rate:

$$\boxed{\gamma \; = \; -\ln \rho_1}$$

### 4.9.2 Recipe for the Average Kinetic Energy $\langle K \rangle$

1. Generate embeddings: $q_t = f(\boldsymbol{v}[t : t + N])$.
2. Compute the semantic velocity: $\dot{q}_t \approx q_{t+1} - q_t$.
3. Instantaneous kinetic energy: $K_t = \frac{1}{2} m |\dot{q}_t|_g^2$.

4. Average: $\langle K \rangle_{\text{measured}} = \frac{1}{n}\sum_{t=1}^{n} K_t$.
5. Estimate $\gamma$
6. Adjust for Brownian motion:

$$\boxed{\langle K \rangle = \langle K \rangle_{\text{measured}} \cdot \mathcal{F}(\gamma)}$$

where $\mathcal{F}(\gamma)$ comes from the theory of Brownian motion, and is:

$$\mathcal{F}(\gamma) = \frac{\gamma^2}{1 - \gamma + \exp(-\gamma)}$$

### 4.10 Temperature $T$

Equipartition gives a practical *thermometer*: in equilibrium, each of the dd quadratic velocity modes carries $\frac{1}{2}T$ of energy, so $T = \frac{2}{d}\langle K \rangle$. Because sampled velocities are finite-difference averages, the raw estimate is biased low by a known factor $F(\gamma\tau)$; divide by $F$ (with $\tau = 1$) to correct it. Once $T$ is fixed, all other thermodynamic quantities inherit a consistent numerical gauge. The formula is exact only if the system is canonical, inertia is constant, and fast modes are well sampled—otherwise $T$ becomes an effective, diagnostic temperature.

### 4.11 Internal Energy $U$

The internal energy—*average semantic energy*—of the cloud of embedding vectors is:

$$U = \langle H \rangle = -\frac{\partial}{\partial \beta} \log Z(\beta).$$

From the factorized $Z(\beta)$:

$$\log Z(\beta) = \frac{d}{2}\log(2\pi m) - \frac{d}{2}\log\beta + \log\left(\int dq\, e^{-\beta V(q)}\right),$$

so:

$$U = \frac{d}{2\beta} + \frac{\int dq\, e^{-\beta V(q)} V(q)}{\int dq\, e^{-\beta V(q)}} = \langle K \rangle + \langle V \rangle.$$

This recovers equipartition, which will prove useful later: $\langle K \rangle = \frac{d}{2}T$.

On the other hand, the average potential energy $\langle V \rangle$ depends on the shape of $V(q)$ and the temperature. At low $T$, $\langle V \rangle$ approaches the global minimum of $V(q)$; at high $T$, it flattens toward the average over $\mathcal{E}$.

This allows us to *measure semantic temperature* directly from observed kinetic energy: a text with high $\|\dot{q}(t)\|$ variance is "hot"; one that stays near a topic center is "cold".

#### *4.11.1 Recipe for $U(T)$*

1. Compute $\langle V \rangle$ (Monte Carlo or parametrization).
2. Compute $\langle K \rangle$.
3. Add:

$$\boxed{U(T) = \langle K \rangle + \langle V \rangle}$$

### 4.12 Entropy $S$

The *Gibbs entropy* quantifies the uncertainty or diversity of semantic states of the ensemble. It is defined as the expectation of the negative log-density:

$$S = -\int dq\, dp\, \rho(q,p) \log \rho(q,p), \quad \rho(q,p) = \frac{1}{Z} e^{-\beta H}.$$

Then, we can write the entropy of the cloud of semantic embeddings as:

$$S = \beta \langle H \rangle + \log Z = \frac{\langle H \rangle - F}{T}.$$

High entropy corresponds to *semantic diversity*—a text that explores many topics or styles. Low entropy indicates *focus or redundancy*, such as repetitive reasoning or narrow discourse. This makes entropy a natural metric for analyzing genre, authorial style, or model behavior.

#### *4.12.1 Formula for $S(T)$*

$$\boxed{S(T) = \frac{U(T) - F(T)}{T}}$$

### 4.13 Semantic Pressure $P$

Building on the definition of $\mathcal{V}_{\text{sem}}$, we define *semantic pressure* $P$ as the thermodynamic conjugate of volume in the canonical ensemble. It quantifies the tendency of the semantic system to expand its scope of meaning in response to confinement.

In the canonical ensemble, the partition function $Z(\beta, \mathcal{V}_{\text{sem}})$ depends both on the inverse temperature $\beta = 1/T$ and the accessible semantic volume. The semantic pressure is then given by:

$$P(\beta) = \frac{1}{\beta} \frac{\partial}{\partial \mathcal{V}_{\text{sem}}} \log Z(\beta, \mathcal{V}_{\text{sem}}).$$

This measures how sensitive the system's free energy is to changes in the available semantic space. A high $P$ indicates strong resistance to confinement—a "drive" to explore new meanings—while a low $P$ suggests contentment within a limited conceptual domain.

For a free semantic particle (i.e., $V(q) = 0$) in $d$-dimensions with Euclidean metric, the partition function factorizes as:

$$Z = \frac{\mathcal{V}_{\text{sem}}}{\lambda_B^d} \implies \log Z = \log \mathcal{V}_{\text{sem}} + \text{const},$$

so:

$$P = \frac{1}{\beta \mathcal{V}_{\text{sem}}} = \frac{T}{\mathcal{V}_{\text{sem}}}.$$

This gives the *semantic ideal gas law*:

$$P \mathcal{V}_{\text{sem}} = T.$$

High $P$: resistance to confinement (creative tension); low $P$: stagnation.

#### 4.13.1 Formula for $P(T)$

Empirically, the pressure of the gas $P$ of *semantic embeddings* can be estimated as:

$$\boxed{P(T) \approx \frac{T}{\mathcal{V}_{\text{sem}}}}$$

where $T = \frac{2}{d}\langle K \rangle$ is the semantic temperature and $V_{\text{sem}}$ is derived from the support of $q(t)$. Applications include detecting narrative build-up (rising $P$) or diagnosing stagnation (low $P$ despite high $T$).

Thus, *semantic pressure* completes the core thermodynamic triad—$T$, $S$, $P$—and enables a richer analysis of linguistic dynamics as a driven, expansive process.

### 4.14 Specific Heat $C_V$

The specific heat at constant volume, denoted $C_V$, is a fundamental thermodynamic quantity that measures the system's ability to absorb energy in response to a change in temperature. In physical systems, it characterizes thermal inertia; in our framework, it quantifies the resistance of a semantic system to changes in agitation (temperature).

In the *gas of semantic embeddings*, we define $C_V$ as the rate of change of the average energy wrt. temperature:

$$C_V = \frac{\partial \langle H \rangle}{\partial T} = \frac{1}{T^2} \frac{\partial^2}{\partial \beta^2} \log Z(\beta) = \frac{\text{Var}(H)}{T^2}.$$

This is a key result in *Statistical Mechanics*: the specific heat is proportional to the fluctuations in energy.

As for interpretation, a high $C_V$ suggests that the system can absorb large changes in temperature with minimal disruption to its average energy — it is *thermally stable*. On the other hand, a low $C_V$ suggests that the system is sensitive to temperature changes — small increases in $T$ cause large increases in $\langle H \rangle$, indicating *semantic fragility*. In linguistic terms, a coherent, well-structured text (e.g., a logical argument) may exhibit high $C_V$: it resists thermal agitation and maintains stability even as $T$ increases.

#### 4.14.1 Recipe for $C_V(T)$

1. Compute $\text{Var}(H) = \langle H^2 \rangle - \langle H \rangle^2$.
2. Compute:

$$\boxed{C_V(T) = \frac{\text{Var}(H)}{T^2}}$$

## 5 Hypotheses

### 5.1 Ergodic Hypothesis

**Time averages along a single trajectory equal ensemble averages over phase space.**

In physics, the long-time trajectory of a system explores all accessible regions of phase space uniformly, so the average of a quantity over time equals its average over all possible states. Similarly, in semantics, we assume that the evolution $(q(t), p(t))$ of a single long text (e.g., a novel) samples the full distribution of semantic states characteristic of its genre, author, or theme.

This is crucial for empirical work: it allows us to treat one book as a proxy for the "statistical behavior" of a writer or genre. Importantly, real texts may violate ergodicity (e.g., narratives have irreversible arcs, authors shift style), suggesting *non-equilibrium statistical mechanics* may be more appropriate in some cases.

### 5.2 Equal A-Priori Probability Hypothesis

**In equilibrium, all accessible microstates consistent with the system's energy and constraints are equally probable.**

In physics, for an isolated system in equilibrium, the probability density $\rho(q, p)$ is uniform over the energy shell $H(q, p) = E$. Equivalently, in semantics, over a long text or corpus in a "stationary" semantic regime (e.g., consistent topic or style), all meaning states that are semantically coherent and dynamically accessible should be equally likely under the model.

This justifies using the *microcanonical ensemble*, where entropy is defined as:

$$S = k_B \log \Omega$$

with $\Omega$ the volume of phase space occupied by states at fixed energy.

### 5.3 Continuum Semantic Trajectory Hypothesis

**The trajectory of the semantic vector can be treated as a partial observation of an underlying continuous trajectory.**

By interpolating over time $t$ the discrete embeddings $q_t = f(\mathbf{v}[t : t + N])$ obtained in a sliding window fashion, we get a smooth curve $q(t) : \mathbb{R}^+ \to \mathcal{E}$. The assumption that this *lifting* can be done without changing the trajectory in a meaningful way opens the door to computing the *momentum* of the particle, which in turn will unlock all the operators and functionals used in *Lagrangian mechanics.*

### 5.4 Equilibrium Hypothesis

**We assume that the gas we are studying is at equilibrium**. In principle, equilibrium occurs when the distribution of embeddings $\rho$ of any (reasonably) small portion of the text is similar to the distribution of the entire corpus.

This hypothesis may not hold in general, and it's an interesting question whether it may be relaxed. In any case, to improve the stability, it should be a good idea to split the text based on meaning and study each chunk individually.

### 5.5 Constant Inertia Hypothesis

**The mass of the semantic particle is constant.** This hypothesis is necessary to study the system with the tools of *Statistical Mechanics*. That aside, the value of mass is not defined in this framework, as all we can measure is velocity $\dot{q}$ and the force per unit mass $\frac{1}{m}\nabla V(q)$.

## 6 Other Thoughts

### 6.1 Diffusion and Stochastic Dynamics

The concept of noise is highly relevant in this context for two main reasons. First, natural language has a significant level of stochasticity. Second, LLMs use noise to make the output text more *creative*, *diverse*, and perhaps *realistic*.

### 6.2 Why Do LLMs Get Stuck?

We now have a framework to study the phenomenon of LLMs repeating periodically the same output token over and over again. If we add more noise to the output, it tends to do that less often, suggesting a notion of *potential well* and *kinetic energy*.

#### *6.2.1 Key Idea #6: Critical Temperature $T_{crit}$*

We propose that by analyzing thermodynamic behavior before looping occurs, we can estimate a safety threshold $T_{\text{crit}}$.

1. Study $K_{\text{avg}}(T^{(model)})$ from generated text.
2. From looping examples, estimate well depth $\hat{V}$ and kinetic energy $\hat{K}$ at the critical point.
3. Set $K_{\text{crit}} = \hat{V}$.
4. Invert to get:
$$T_{\text{crit}} = T^{(model)}(K_{\text{crit}})$$

Above $T_{\text{crit}}$, kinetic energy overcomes potential barriers, breaking loops.

## 7 Conclusion

Semantic Dynamics provides a principled framework for analyzing LLM behavior through statistical mechanics. By mapping token sequences to trajectories in a latent energy landscape, we derive thermodynamic quantities that diagnose and mitigate degenerate generation. We estimate a critical temperature $T_{\text{crit}}$, above which models escape semantic wells. This offers a new path toward more coherent, diverse, and stable language model outputs.

The analogy of a "semantic particle" moving through meaning space transforms abstract language into a physical-like system, enabling diagnosis via temperature, pressure, and entropy.

Future work includes non-equilibrium extensions, curvature-aware metrics, and applications to style transfer and cognitive modeling.

## 8 Other Ideas

**Quasi-static Approximation:** In this framework, we assume a certain notion of equilibrium within the corpus of text. However, we might be able to study **Canonical Transformations:** Apply $(q, p) \mapsto (Q, P)$ preserving symplectic structure. Useful for style transfer or paraphrasing.

**Semantic Potential Landscapes:** Map long texts into energy landscapes; identify topic clusters (wells) and transitions (barriers).

**Semantic Turbulence:** Analyze power spectrum of $p(t)$ or $\dot{q}(t)$; high frequencies may indicate cognitive load or emotional intensity.