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 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 like looping or stuck states. Our approach provides a practical diagnostic tool to estimate when a model is trapped in a semantic potential well and how to escape it by tuning an effective temperature, directly addressing repetition and periodic token cycles in LLMs.

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 centered around position t. From now on, the dependency on $\boldsymbol{v}$ is implied but omitted.

1.1 Key Idea #1: Continuum Semantic Trajectory Hypothesis

We treat the discrete sequence of embeddings $\boldsymbol{q} = (q_t)$ as partial observations of an underlying continuous trajectory $q(t) : \mathbb{R}^+ \to \mathcal{E}$. By interpolating the discrete embeddings $q_t = f(\boldsymbol{v}[t : t+N])$ obtained via a sliding window, we obtain a smooth curve $q(t)$.

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 velocity leads to the *momentum*:

$$p(t) = \frac{\partial L}{\partial \dot{q}}(q(t), \dot{q}(t), t),$$

where the Lagrangian L implicitly depends on a notion of *semantic inertia* m. While m may vary (e.g., scientific texts resist change more than poetic ones), we assume constant inertia for simplicity (see *Constant Inertia Hypothesis*).

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 opens the door to analyzing linguistic dynamics using tools from statistical mechanics—energy, entropy, temperature, and diffusion—as physical-like processes.

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 sensitivity of meaning to context perturbations.

4 Recovering the Thermodynamic Quantities

4.1 Key Idea #3: Canonical Ensemble

With the geometric structure in place, we define physical analogs using the canonical ensemble. Starting from the density $\rho(q)$, we derive all thermodynamic quantities.

4.1.1 Semantic Density $\rho(q)$

The probability density $\rho(q)$ quantifies how frequently regions of $\mathcal{E}$ are occupied. High-density regions correspond to common or coherent meanings; low-density areas represent rare or disfluent constructions. $\rho(q)$ is a proxy for semantic plausibility and defines the potential landscape via:

$$V(q) = -\log \rho(q).$$

4.1.2 Key Idea #4: Estimating $\rho(q)$

To estimate $\rho(q)$ empirically:

1. Compute embeddings: $q_t = f(\boldsymbol{v}[t : t + N])$.
2. (Optional) Apply dimensionality reduction (e.g., PCA).
3. Estimate $\rho(q)$ using KDE, GMM, or k-NN density estimation.

4.2 Dimensionality d

The ambient space is $\mathbb{R}^d$, but dynamics are likely confined to a lower-dimensional manifold $\mathcal{M} \subset \mathbb{R}^{d_{\text{eff}}}$ due to linguistic constraints.

4.2.1 Key Idea #5: Dimensionality Reduction

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 thermodynamic formulas, mitigating the curse of dimensionality. Throughout, d should be interpreted as d_{eff}—the true number of independent semantic degrees of freedom.

4.3 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.

A large $\mathcal{V}_{\text{sem}}$ indicates broad exploration; a small $\mathcal{V}_{\text{sem}}$ suggests focused discourse.

4.3.1 Algorithm for $\mathcal{V}_{\text{sem}}$

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

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

4.4 Potential Energy $V(q)$

$V(q)$ is the semantic landscape guiding dynamics. Under the Gibbs-Boltzmann hypothesis and equilibrium:

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

with $\beta = 1/T$, and $k_B = 1$ (natural units).

4.4.1 Recipe for $V(q)$

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

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

4.5 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:

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

4.6 Partition Function $Z(\beta)$

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 $\lambda_B = \sqrt{\beta/(2\pi m)}$:

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

4.7 Free Energy $F(\beta)$

The Helmholtz free energy is:

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

It balances semantic coherence (low V) and diversity (high S).

4.7.1 Recipe for $F(T)$

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

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

4.8 Internal Energy U

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

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: $\langle K \rangle = \frac{d}{2}T$.

4.8.1 Recipe for $U(T)$

1. Compute $\langle V \rangle$ (Monte Carlo or parametrization).
2. Compute $\langle K \rangle = \frac{d}{2}T$.
3. Add:

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

4.9 Entropy S

Gibbs entropy:

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

Then:

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

High S = semantic diversity; low S = redundancy.

4.9.1 Formula for $S(T)$

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

4.10 Semantic Pressure P

Conjugate to $\mathcal{V}_{\text{sem}}$:

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

For a free particle ($V = 0$):

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

so:

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

This gives the *semantic ideal gas law*:

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

4.10.1 Formula for $P(T)$

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

High P: resistance to confinement; low P: stagnation.

4.11 Specific Heat C_V

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

High C_V: thermally stable (e.g., logical argument); low C_V: fragile.

4.11.1 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 Locking Everything into Place

We can determine T by measuring one quantity.

5.1 Method 1: Measuring Average Kinetic Energy

1. Generate embeddings: $q_t = f(\boldsymbol{v}[t : t + N])$.
2. Compute 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: $\langle K \rangle = \frac{1}{N} \sum_{t=1}^{N} K_t$.
5. Compute temperature:

$$\boxed{T = \frac{2}{d} \langle K \rangle}$$

Note: Sensitive to derivative approximation, especially under Brownian dynamics.

5.2 Method 2: Measuring Entropy

5.2.1 Key Idea #6: Locking via Entropy

Estimate entropy via Lempel-Ziv complexity: $S(T) \approx S_{\mathrm{LZ}}$. Invert $S(T)$ to get $\hat{T} = T(S_{\mathrm{LZ}})$, then plug into other quantities.

1. Compute $S(T)$ as above.
2. Invert to get $T(S)$.
3. Estimate S_{LZ} from token sequence.
4. Set $\hat{T} = T(S_{\mathrm{LZ}})$.
5. Use $\hat{T}$ to compute all other quantities.

6 Hypotheses

6.1 Ergodic Hypothesis

Time averages equal ensemble averages. A long text samples the full distribution of semantic states for its genre or author. Crucial for empirical validity, though narratives may violate ergodicity.

6.2 Equal A-Priori Probability Hypothesis

In equilibrium, all accessible semantic states at fixed energy are equally probable. Justifies microcanonical ensemble:

$$S = \log \Omega, \quad \Omega = \text{phase space volume at } H = E.$$

6.3 Continuum Semantic Trajectory Hypothesis

Discrete embeddings are samples from a continuous $q(t)$. Enables calculus-based mechanics.

6.4 Equilibrium Hypothesis

The embedding distribution $\rho(q)$ is stationary across the text. May be relaxed; consider analyzing chunks separately.

6.5 Constant Inertia Hypothesis

Semantic mass m is constant. Necessary for standard statistical mechanics.

6.6 Hypothesis on Entropy

$S(T) \approx S_{\mathrm{LZ}}$. An approximation; future work may improve this bridge.

7 Other Thoughts

7.1 Diffusion and Stochastic Dynamics

Model noise via stochastic differential equations:

$$dq(t) = \dot{q}(t)\,dt, \quad d\dot{q}(t) = F(q, \dot{q})\,dt + \sigma\,dW_t,$$

where $F = -\nabla V$, and W_t is Wiener noise. Leads to Fokker–Planck equation for $\rho(q, \dot{q}, t)$.

7.2 Why Do LLMs Tend to Get Stuck?

LLMs repeat tokens when trapped in potential wells. Adding noise reduces looping, suggesting a kinetic escape mechanism.

7.2.1 Key Idea #7: Critical Temperature T_{crit}

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

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

8 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_{crit}, above which models escape semantic wells. This offers a new path toward more coherent, diverse, and stable language model outputs.

9 Other Ideas

- **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.