Semantic Guage Theory ™ of AI Hallucination

Patent Pending, USPTO Filing Date June 23, 2025

By Ben Mancke

Abstract:

Construction of a canonical set semantic translations as a Cayley graph from Zipfian ranked points in LLM embedding space. Canonical translations are used to build a local lie algebra in a neighborhood around a point in embedding space. A Taylor polynomial centered around an LLM's output approximates a composition of each LLM transformation layers. The Jacobian of this function is projected onto the local Lie algebra basis around the input vector is constructed to create a mathematically novel complex order covariant derivative of Generative transformation in a semantic direction. The components satisfy a gauge theory of AI Hallucination and explain it phenomenologically as a non-entropic local property of manifold curvature. Intended use in study and suppression of LLM hallucination, implementation of LLM self-learning architecture, and reduced compute cost. All of this is applied to a static pretrained embedding space and requires no additional LLM training.

Table of Content

1. Related Topics and readability
2. Overview of conventional LLM
    2.1. Key Components:
    2.2. Example
3. Semantic Directionalization
    3.1. Overview
    3.2. Construction of Canonical Semantic Translations
    3.3. Construction of Local Lie Algebra Basis
4. LLM Layer Deformation
    4.1. LLM Layer Transformations
    4.2. Example using wheel rotation
    4.3. Layer Deformation Composition
5. Differential Geometry
    5.1. Principal concept
    5.2. Analog of dx
    5.3. Analog of dy
    5.4. The Derivative of Generative Deformation with respect to Semantics
    5.5. Parallel Transport
    5.6. Order Extension
6. Bracket Residual
7. Semantic Gauge Theory
8. Phenomenology

1. Related Topics and readability.

To get though the material and fully appreciate the technical details would require some level of pedestrian familiarity with Group theory, Jacobians, Differential Geometry, Lie Algebras, Zipf distributions, Locally Testable Codes (LTCs), and the PageRank algorithm, the Special Orthogonal Goup SO(n).

2 Overview of conventional LLM

A large language model (LLM) is a kind of neural network trained to predict the next word in a sentence. It learns to do this by ingesting vast amounts of text and updating its internal parameters to minimize prediction error. Over time, this process builds a deep statistical understanding of language.

2.1 Key Components:

Tokenization: Text is broken into tokens (e.g., words or word pieces like "play" and "ing").

Embedding Layer: Each token is mapped to a high-dimensional vector. This is the embedding: a geometric representation of meaning. Similar meanings cluster together (e.g., cat and dog).

Transformer Architecture: Uses attention mechanisms to model relationships between words, regardless of distance. Deep layers process how each word relates to others in the context of the sentence. Meaning is progressively deformed as each transformer layer applies a nonlinear context-dependent operation to the token embeddings.

Output Layer: For each token, predicts the next most likely token using a probability distribution over the vocabulary.

Training: The model is trained on billions of examples using a loss function (like cross-entropy) and gradient descent.

2.2 Example

Input: "Paris is the capital of ___"

How the LLM handles it:

1. Tokenizes: → ["Paris", "is", "the", "capital", "of"]
2. Embeds each token as a point in a vector space.
3. Uses attention to determine that "Paris" heavily influences the missing word.
4. Predicts the most likely next token: → "France"

Why it works:

Because the model has seen thousands of examples like: "Paris is in France", "The capital of France is Paris", "Berlin is the capital of Germany". It learns statistical patterns of how words relate in context, encoded geometrically in the embedding space.

3 Semantic Directionalization

3.1 Overview

Vector addition in embedding space models discrete, composable semantic transformations between concepts. This allows the construction of a Canonical Cayley Diagram of semantic transformations using a combination of Zipfian frequency and PageRank algorithm rankings and filtering.

This structure is used to create a local semantic Cayley patch and Lie Algebra basis. This takes the semantic changes as jumping from one point to another (like man → woman), and turns it into a continuous flow that, most importantly, Semantically Directionalizes a local region of the LLM embedding space.

3.2 Construction of Canonical Semantic Translations

Below is what amounts to an algorithm to construct a relatively low compute (especially in LLM terms) algorithm for building a one-and-done reuseable set of Canonical Semantic Translations.

1. Establish a Translation Radix $r \in \mathbb{N}$ of the canonical translation set
2. Select Zipf-Ranked Vocabulary
   a. Extract the top $\frac{r^6}{2}$ tokens or phrases from the LLM's vocabulary.
   b. Prioritize high-frequency, semantically rich terms (e.g., “tense”, “plurality”, “negation”, “opposition”).
3. Generate Pairwise Semantic Transitions
   a. Construct approximately r^6 candidate token pairs of semantically linked terms, such as:
      i. Tense: bite → bit
      ii. Plurality: person → people
      iii. Antonyms: hot → cold
   b. Compute difference vectors: $\Delta_{ij} = v_i - v_j$
   c. Normalize all Δ_{ij}
4. Cluster into Canonical Directions
   a. Apply k-means, k-medoids, or spectral clustering to the set of Δ vectors.
   b. Extract cluster centers μ_k as prototype transformations.
   c. Each center μ_k is a potential canonical transformation.
5. Semantic Validation
   a. For each cluster center μ_k, randomly select $\leq \frac{r^2}{2}$ token pairs from the cluster.
   b. Use the LLM to evaluate semantic coherence of each transformation.

c. Retain clusters whose test pairs exhibit consistent semantic behavior (e.g., “singular → plural” remains stable across examples).

d. Keep only the first r^5 validated transformations.

6. Graph-Based Filtering

a. Build a semantic transition graph $G = (V, E)$:

i. Nodes = token embeddings

ii. Edges = validated μ_k transformations

b. Run PageRank on this graph to identify topologically central and compositionally robust transformations.

c. Select the top r^4 canonical vectors $\{\mu_k\}_{k=1}^{r^4} = T$ that participate in closed compositional paths on the graph (indicative of structural reuse and symmetry).

d. Retain r^3 where for each μ_i there exists $\{\mu_{ij}\}$, $\{\mu_{ik}\}$ such that:

i. $\mu_i \notin \{\mu_{ij}\},\ \mu_i \notin \{\mu_{ik}\}$

ii. $\sum_j \mu_{ij} = \mu_j$ and $\sum_k \mu_{ik} = \mu_k$

iii. $\frac{\mu_i}{||\mu_i||} \approx -\frac{\mu_j}{||\mu_j||}$ and $\mu_i \cdot \mu_h \approx 0$

iv. Here, we define the notion of sufficient orthogonality and normalized equality above in terms of cosine similarity

v. Every Canonical Translation μ_i has a distributed orthogonal and reverse

3.3 Construction of Local Lie Algebra Basis

1. Begin with a point v in embedding space and the canonical transformation set T

a. Use cosine similarity to select a subset $T_{local} \subset T$ such that:

i. $\mu_k \in T_{local} \Leftrightarrow \left(\frac{\mu_k \cdot v}{||\mu_k||\,||\mu_k||}\right) < \theta$

ii. Where θ is sufficiently small (~15°) to produce a local semantic basis

2. Build a directed graph $G_{patch} = (V, E)$ where:

a. $V = \{v_i\}$ contains embeddings points

b. $E \subseteq T_{local}$ contains semantic translations

c. $v -_{\mu_i}\rightarrow v_1 -_{\mu_j}\rightarrow v_2$ for $\mu_i \in E$ and $v_i \in V$

3. Apply Local Testable Codes (LTC)

a. Validate that transitions between nodes in G_{patch} form coherent semantic trajectories.

b. Use LLM-assisted local consistency checks (e.g., triangle closure, word pair reversibility) to prune invalid paths.

4. Define Local Lie Algebra Basis

a. Use LTC filtered subset of E as generators of a local Lie algebra such that for a point v in the embedding space:

i. $g_v = Span\{X_i\}$, where $\{X_i\}$ is Graham Schmidt orthonormalized subset of $\{\mu_k\}$ and each X_i has a distributed orthogonal and reverse in g_v

ii. This constructs an orthonormalized local Lie basis that spans a semantic neighborhood around v

b. From the above we can state that g_v
   i. Is cosign similarity invariant under rotations in SO(n)
   ii. $\cup_{R\in SO(n)}\ R\ \cdot \text{span}(\{X_i\}) = \mathbb{R}^n$ - Rotation linked Lie groups span embedding space
   iii. There exists a smooth local semantic manifold around every point in embedding space whose span is g

4 LLM Layer Deformation

4.1 Layer Transformations

When an input vector from the embedding space passes through the transformer layers, it is transformed and sent to the next layer. The layer transformations acting on the input are then projected back to embedding space. The composition of each transformation is interpolated near the output projection to define a locally continuous function that maps the input to the output

4.2 Example

Suppose $E\colon\ \mathbb{R}^n \rightarrow \mathbb{R}^n$ is the canonical embedding and $\Phi_{context}\colon\ \mathbb{R}^n \rightarrow \mathbb{R}^n$ is a spatial deformation. This gives a new transformed space Φ_{context}(E) where the relative distances and directions between meanings are altered.

- “woman” might drift toward "mother" or "wife"
- “man” might drift toward "leader" or "hero"

This results in a transformation like: $\Phi_{bias}(woman) = woman + \Delta_{\ cultural\ exectations}$

4.3.1 Layer Deformation Composition

In modern LLMs meaning is deformed as each transformer layer applies a nonlinear context-dependent operation denoted f_i for the i th transformation layer. The total deformation F is the composition of each f_i:

$$f_n \circ f_{n-1} \circ \ \ldots . \ \circ f_1 \ = \ F$$

4.3.2 Composite Approximation

To analyze this deformation in geometric terms, the model’s hidden state vector h is projected back to embedding space via the output embedding matrix M_{out} to get the output vector.

$$M_{out} \cdot h \ = \ v_{out}$$

A 4th degree Talor polynomial centered around v_{out} is then used to approximate F

$$\sum_{|m|\geq 0} \frac{1}{m!} D^m F(v_{out}) \cdot (x - v_{out})^m \approx \Phi(x)$$

Where:

$$\sum_i m_i = \ |m|, \ \prod_i mi = m!, \ \prod_i (x_i - v_i)^{m_i} = (x - v)^m$$

$$\prod_j \quad \frac{\partial^{m_j}}{\partial x_i^{m_j}} = D^m$$

Similarly compose each f_i^{-1} and interpolate at v_{in} to get Φ^{-1}:

$$f_1^{-1} \circ \ f_2^{-1} \circ \ldots \circ \ f_n^{-1} = F^{-1} \approx \Phi^{-1}$$

5 Differential Geometry

5.1 Principal concept

The Function Φ transforms a region around the input vector v_{in}, deforming the space around as v_{in} is mapped to v_{out}. Specifically, measuring how much and in what direction LLM layer transformation locally changes embedding space with respect to small changes in semantics

5.2 Analog of dx

Previously it was explained and justified constructively that local areas of embedding space can be treated as a local manifold analogous to the conventional x-axis. Now that is used to build a differential operator.

Let G be a group of semantic operations on a local Cayley Patch and let the Lie algebra g associated with the group G be a real vector space:

$$g = \ \text{span}\,\{X_1, X_2, \ldots, X_n\}$$

Where:

- $X_i(x) = \left[\frac{d}{dx}\ \right]_{t=0} (\exp(tX) \cdot x) = \frac{dX_i}{dx}$
- Each X_i is a generator: a direction of infinitesimal transformation in semantic space (Cayley graph).
- These generators form a basis of the Lie algebra.
- These basis elements are implemented as matrices (square matrices of size equal to the embedding dimension).

- They satisfy the Lie bracket: $[X_i, X_j] = X_i X_j - X_j X_i = \sum_k c_{ij}^k X_k$

Where c_{ij}^k are the Lie algebra structure constants.

5.3 Analog of dy

If $x = v_{in}$ is the LLM's input vector in embedding space, then a transformation layer deformation $\Phi(x) \approx v_{out}$ in the same space. The Jacobian matrix $J_\Phi(x)$ describes how this transformation changes with respect to small changes in x . It's defined by the partial derivatives of Φ :

$$[J\Phi(x)]_{ij} = \frac{\partial \Phi_i(x)}{\partial x_j}$$

Where each entry tells you how the i th output dimension of the LLM Layer transformed vector $\Phi(x)$ changes with respect to the j th input dimension:

$$\Phi(x + \epsilon) \approx \Phi(x) + J\Phi(x) \cdot \epsilon$$

5.4 The Derivative of Generative Deformation with respect to Semantics

From the local Lie algebra, we have a differential for sematic change in the embedding space. From the Jacobian we have a differential for LLM layer Transformation change in the embedding space. Let us define the projection of the Jacobian onto the Lie Algebra Basis to get a derivative of Generative transformation in a semantic direction.

$$\frac{d\Phi}{dX_i} \overset{\text{def}}{=} \nabla\Phi \cdot X_i = \left\langle \frac{d\Phi}{dx}, X_i \right\rangle_F$$

Where:

- X_i is the i -th Lie algebra generator (a directional vector field)
- $\nabla\Phi$ is the Jacobian of Φ
- The inner product gives the component of the deformation in the X_i direction

5.5 Parallel Transport

So far, we have worked with a Taylor Polynomial interpolant of the layer composite transformation around the model's output v_{out} in the embedding space. But we can pass the local Lie algebra basis through the layers along with the input, and track the deformation at each layer

Let $\text{span}\{X_i\} = g$ be the local lie basis around v_{in}

Let $f_k(v_{k-1}) = v_k$ denote the k -th layer acting on the output of the previous layer

Let $X_i^{(k)} \;=\; \frac{f\left(X_i^{(k-1)}\right)}{\left\|f\left(X_i^{(k-1)}\right)\right\|}$ for $k > 1$ and $X_i^{(1)} \;=\; X_i$ (Layer-wise normalization)

Let $\left\{X_{iGS}^{(k)}\right\} = \left\{f_k^{GS}\left(X_i^{(K-1)}\right)\right\}$ denote the Grahm Schmidt orthonormalization of $\left\{f_k\left(X_i^{(k-1)}\right)\right\}$

Let $\Phi_k \;\approx\; f_k$ be a local 4th degree Taylor polynomial interpolant centered at v_k

From the above, have a lie basis that serves as a fiber bundle that is transported through the layers. From the above know facts, it is necessarily the case that the covariant derivatives exist in each layer.

$$\frac{d\Phi_k}{dX_i^{(k)}} = \left\langle \frac{d\Phi_k}{dx}, X_i^{(k)} \right\rangle \text{ and } \frac{d\Phi_k}{dX_{iGS}^{(k)}} = \left\langle \frac{d\Phi_k}{dx}, X_{iGS}^{(k)} \right\rangle$$

It should also be noted that instead of using the parallel transport, one could also use the previous composition and back projection method to work in the embedding space.

5.6 Order Extension

A covariant derivative like this has been seen in applications of physics. But the construction of this is novel in terms of its application. For AI research this is may well be a breakthrough already.

Extending the order of the Jacobian to the complex numbers produces an object that, while prepared from many pre-existing ingredients, is itself an entirely novel cocktail not just in application, but novel in terms of pure mathematics.

Recall the conventional h- difference quotient limit definition of a function f: $\mathbb{R}^n \to \mathbb{R}^n$ from undergraduate calculus:

$$\lim_{h\to 0} \frac{f(x+h) - f(x)}{h} = f'(x)$$

Define a complex order derivative from the Grünwald–Letnikov (GL) Derivative:

$$\lim_{h\to 0} \frac{1}{h^{\alpha}} \sum_{n=0}^{\infty} (-1)^n \; \binom{\alpha}{n} f(x - nh) \;=\; D^{\alpha} f(x)$$

Where $\alpha \in \mathbb{C}$, $h \in \mathbb{R}$ and $\binom{\alpha}{n} \;=\; \Gamma(\alpha)$

Define a complex order partial derivative:

$$\lim_{h\to 0} \frac{1}{h^{\alpha}} \sum_{n=0}^{\infty} (-1)^n \; \binom{\alpha}{n} f\left(x - nhe_j\right) \;=\; \frac{\partial^{\alpha} \Phi_i}{\partial x_j^{\alpha}}$$

Where e_j is a unit vector in the j -th direction

Define a new complex order Jacobian:

$$[J_{\Phi}^{\alpha}(x)]_{ij} \ = \ \frac{\partial^{\alpha}\Phi_i}{\partial x_j^{\alpha}}$$

Project onto the Lie algebra basis to get the complex order covariant derivative in a semantic direction.

$$\left\langle \frac{d^{\alpha}\Phi}{dx}, X_i \right\rangle = \frac{d^{\alpha}\Phi}{dX_i}$$

Note that only the order of the Jacobian has been modified. The projection is onto the same local Lie Basis.

6 Bracket Residual

6.1 Claim: Structure is preserved under real number order covariant differentiation

Let Φ : : $\mathbb{R}^n \rightarrow \mathbb{R}^n$ be local interpolant of a Layer Transformation of the embedding space near the vector x .

Let X_i and X_j be elements of the local Lie algebra basis

Let $\alpha \in \mathbb{R}^+$.

Then:

$$\left[\frac{d^{\alpha}}{dX_i}, \frac{d^{\alpha}}{dX_j}\right]\Phi(x) \ = \ \sum_k c_{ij}^k \frac{d^{\alpha}\Phi}{dX_k}(x)$$

Where:

- $\frac{d^{\alpha}}{dX_i}$ is the complex order derivative along the Lie generator X_i
- c_{ij}^k are the Lie algebra structure constants
- The local lie group is a submanifold of $\mathbb{R}^n$ implying sufficient smoothness
- Φ is 4th degree polynomial, implying sufficient smoothness
- Linearity of differential operator holds under the real order partial derivative definition, applied consistently across the lie group manifold.

This implies Lie group closure of the real number order operators impact only scaling of flows in the direction of X_i and does not change structure constant orientation.

- $\left[\frac{d^{\alpha}}{dX_i}, \frac{d^{\alpha}}{dX_j}\right]\Phi(x) \;=\; \frac{d^{\alpha}}{dX_i}\left(\frac{d^{\alpha}\Phi}{dX_j}(x)\right) - \frac{d^{\alpha}}{dX_j}\left(\frac{d^{\alpha}\Phi}{dX_i}(x)\right)$
- $\left[\frac{d^{\alpha}}{dX_i}, \frac{d^{\alpha}}{dX_j}\right] \;\propto\; \sum_k \quad c_{ij}^{k}\frac{d^{\alpha}}{dX_k}$

Conclusion: Group operation and structural orientation hold under real number order differentiation. Which makes this a fully developed real order covariant derivative of Generative Layer deformation in a semantic direction. Moreover, this holds for both the full model transformation projected back to the embedding space, as well as the Layer-wise via Parallel Transport. Local curvature can be measured anywhere in the model.

6.2 Claim: Imaginary order differentiation breaks bracket closure

Let all assumptions from Theorem 1 hold but instead let $\alpha \in \mathbb{C} \setminus \mathbb{R}$ and let $\mathrm{Re}(\alpha) = r$ and $\mathrm{Im}(\alpha) \neq 0$. Then:

Recall the limit definition for a complex order GL derivative contains $h^{\alpha} = h^{a-bi}$, which breaks the linearity of differentiation.

$$\lim_{h\to 0} \quad h^{a-bi} = \lim_{h\to 0} \quad h^{a}e^{ib\ln\left(\frac{1}{h}\right)}$$

Then group closure would imply that $\mathrm{im}(\alpha) = 0$

Then we can define a commutor breaking residual such that it cannot be expressed as a linear combination of the local Lie basis

$$\left[\frac{d^{\alpha}}{dX_i}, \frac{d^{\alpha}}{dX_j}\right]\Phi(x) - \left[\frac{d^{\Re(\alpha)}}{dX_i}, \frac{d^{\Re(\alpha)}}{dX_j}\right]\Phi(x) \;=\; \mathcal{E}_{ij}(x)$$

- $\mathcal{E}$ correspond to the imaginary portion of the complex order derivative
- $\mathcal{E}$ vanishes when $\mathrm{Im}(\alpha) = 0$

- ζ denote local imaginary curvature:

$$\left\|\left[\frac{d^{\Re(\alpha)+\beta i}}{dX_i}, \frac{d^{\Re(\alpha)+\beta i}}{dX_j}\right]\Phi(x) - \left[\frac{d^{\Re(\alpha)}}{dX_i}, \frac{d^{\Re(\alpha)}}{dX_j}\right]\Phi(x)\right\| \;=\; \left\|\mathcal{E}_{ij}(x)\right\| \;=\; H(\beta)$$

- Let T_{ζ} denote total imaginary curvature

$$\int_{-t}^{t} H(\beta)dt$$

7 Semantic Gauge Theory

7.1 Fiber Bundle

- Assertion: Embedding space is treated as a differentiable base manifold, with local patches equipped with Lie algebraic structure.
- Justified in: Section 2.1 and 2.2
    - Canonical translations are selected to span local neighborhoods.
    - Gram-Schmidt orthonormalization ensures tangent space structure.

7.2 Lie Group and Lie Algebra

- Assertion: Local transformations form a group under matrix composition and vector field brackets.
- Justified in: Section 2.3 and 2.4
    - Canonical translations are shown to admit a group closure condition under commutation.
    - Each basis element has an identifiable orthogonal and reverse under cosine similarity (distributed orthogonality).

7.3 Connection Form and Covariant Derivative

- Assertion: A connection is induced via directional derivatives over the Lie basis; deviations arise from curvature.
- Justified in: Section 3.1 and 3.2
    - GL-derived complex-order differential operator defines projection and directional flow.
    - Residual terms from failed bracket closure encode curvature.

7.4 Curvature Tensor

- Assertion: The imaginary component of the complex-order Lie bracket encodes a residual curvature.
- Justified in: Section 3.3 – 3.4
    - Residual term explicitly isolated and shown to be non-zero in general.
    - Linearity failure of the complex-order Jacobian leads to semantic torsion.

7.5 Gauge Invariance and Local Symmetry

- Assertion: Residual curvature is invariant under local orthonormal rotations (SO(n) symmetry).
- Justified in: Section 4.1
    - Lie basis defined to preserve distributed orthogonality and reversibility.
    - Canonical translations allow symmetry-preserving transport.

8 Phenomenology

8.1 Hallucination is not entropy.

It's phenomenologically very much not entropy at all. It's deterministic. It's a local property of a manifold. Specifically, it's a property of strong imaginary curvature in the local geometry, and the Gauge Invariant measures it.

The reason being complex order curvature points in partial semantic directions that only exist locally and momentarily. The larger the imaginary curvature, the greater the divergence from reality. A simple example might be making up a rule for a game with finite well defined rules.

8.2 Experimental Results

Small scale low granularity simulations on toy models easily verify the theory. It's very testable and comports with what the theory predicts. It empirically confirms Commuter non-closure correlates with hallucinated outputs

8.3 Significance

This is a measure a generative hallucination not heuristically but phenomenologically. It tells us hallucination is property of the Layer Transformation curvature on sematic structure, how to measure it, and under what conditions it vanishes. All of this is in terms of a predicable phenomenon of local semantic curvature.

What does it all mean?

- LLM behavior is not statistics at all. It's differential geometry.
- Error can be measurably predicted at any point in any layer
- Any static model can be strategically interpolated with enough compute
- Semantics and meaning are local, and gauge constrained
- Coverage all embedding space by SO(n) linked fiber bundles implies bounded parameterization
- Training coherence corresponds to an LLM's set of Canonical Semantic Translations.
- Canonical transformation set of any AIs embeddings (not just LLMs) extend the gauge theory to all AI regimes
- This mathematical structure takes AI to the next step beyond mere generative prediction. It's the missing link for advancement in alignment, self-improvement, real time correction. And it does this from a phenomenological paradigm.
- Many other paradigm-shifting implications follow immediately from this work.