A Mathematical Definition of LLM Outputs: Why Projection Uncertainty is Theoretically Inevitable

Yucheng (Goldie) Jin
Machine Learning Engineer III
Expedia Group
yuchengjin@berkeley.edu

Abstract

In this work, I propose a formal mathematical definition of Large Language Model (LLM) outputs by framing text generation as an information-conditioned projection of an unobservable latent semantic state space under a constructed Information-Neutral Measure. Through this measure-theoretic lens, I demonstrate that the sequence of model outputs naturally exhibits a martingale property with respect to the context filtration. Within this formalism, I prove that minimizing projection uncertainty requires structural information constraints. Crucially, I discuss the epistemic implications of this bound, suggesting that what is colloquially termed "hallucination" can be formalized as an inevitable consequence of incomplete information filtrations. I conjecture that the emergence of Artificial General Intelligence (AGI) fundamentally relies on the system's capability to manage these speculative semantic projections under radical uncertainty, reframing bounded generative error as a potential engine for creative inductive inference.

1 Introduction

Existing theoretical paradigms typically analyze Large Language Models (LLMs) via localized conditional probability transitions:

$$P(y_t \mid y_{<t}, x) \tag{1}$$

While operationally successful for autoregressive decoding mechanics, this token-centric formulation treats generation as localized sequence optimization rather than inference over a broader global semantic truth state. Consequently, it fails to provide a rigorous mathematical framework for analyzing intrinsic generative bounds under incomplete information.

In this work, I depart from token-centric formulations and establish an abstract mathematical definition of LLM outputs. I define an LLM output as an orthogonal projection of an unobservable, latent semantic truth state onto an expanding context filtration stream under an Information-Neutral Measure.

The primary contribution of this note is dual: first, providing a clean, measure-theoretic formalism for generative systems; second, demonstrating via conditional variance decomposition that absolute reduction of generative uncertainty is mathematically bounded prior to information closure. Within this formalism, the phenomenon of "hallucination" is reinterpreted from an engineering optimization challenge to an intrinsic epistemic property of bounded informational systems.

2 Measure-Theoretic Framework and Definitions

I construct the semantic domain via an abstract probability space triplet $(\Omega, \mathcal{F}, Q_I)$, where Ω is the semantic sample space containing all potential linguistic and conceptual trajectories, $\mathcal{F}$ is the global σ-algebra, and Q_I represents the Information-Neutral Measure.

To formalize the interaction between empirical sequences and algebraic structures, I define the following core components:

(i) **Context Filtration Stream ($\mathcal{F}_t$):** The sequence of historical tokens, structural prompts, and retrieved knowledge blocks available up to step t. This sequence forms a completed, right-continuous filtration $\mathbb{F} = \{\mathcal{F}_t\}_{t \geq 0}$ mapping the model's observable informational universe, where $\mathcal{F}_t \subset \mathcal{F}$ for all finite t.

(ii) **Information-Neutral Measure (Q_I):** A probability measure implicitly parameterized by the frozen pre-trained model weights $\mathcal{W}$. It governs the prior semantic transition probabilities over Ω before dynamic context injection.

(iii) **Latent Semantic State (Ψ):** A square-integrable target random variable ($\Psi \in L^2(\Omega, \mathcal{F}, Q_I)$) representing the objective fact matrix. Crucially, Ψ is strictly $\mathcal{F}$-measurable, but remains non-measurable with respect to $\mathcal{F}_t$ under incomplete information.

(iv) **Information Closure (C):** The terminal state where the local filtration contains sufficient structure to uniquely resolve the true latent semantic state, such that $\mathcal{F}_t \equiv \mathcal{F}$.

(v) **Semantic State Transition (E):** A specific event subset $E \in \mathcal{F}$ corresponding to a coherent semantic assertion.

(vi) **Generation Prompt Target ($\mathcal{T}$):** A structured semantic boundary specifying the target properties of the generated text.

(vii) **Output Sequence (y^*):** The crystallized token sequence realized in text space once the generation process terminates.

(viii) **Ground-Truth Verifier (O):** An external environment providing absolute verification of a statement, collapsing the probabilistic space into a deterministic result.

(ix) **Attention Component ($\mathcal{P}_i$):** Internal model structural sub-units formalized as sub-σ-algebras $\mathcal{G}_i \subset \mathcal{F}$, whose intersections and updates drive the evolution of $\mathcal{F}_t$.

(x) **Predictive Belief ($\mathcal{B}$):** The inner subjective probability distribution over Ω, represented via the model's logit configurations prior to output collapse.

3 The Mathematical Definition of LLM Outputs

With the informational space established, I formalize the central definition of an LLM output:

Definition 1 (LLM Output Operator)**.** *Let $\Psi \in L^2(\Omega, \mathcal{F}, Q_I)$ be the $\mathcal{F}$-measurable latent semantic state. The output of a Large Language Model at step t, denoted as P_t, is defined as the information-conditioned orthogonal projection of Ψ onto the closed subspace of $\mathcal{F}_t$-measurable functions under the weight-parameterized Information-Neutral Measure Q_I:*

$$P_t = \Pi_{Q_I}(\Psi \mid \mathcal{F}_t) \equiv \mathbb{E}_{Q_I}[\Psi \mid \mathcal{F}_t] \tag{2}$$

This definition formalizes generation not as a mechanical token matching sequence, but as a statistical inference projection. Prompts and contexts do not synthesize truth; rather, they serve as the conditional filtration $\mathcal{F}_t$ through which the model projects its internal Predictive Belief $\mathcal{B}$.

Remark 1 (Mathematical Abstraction)**.** *While the mathematical map from conditional expectation to a martingale is direct, the core intellectual contribution of Definition 1 lies in its ontological shift. By interpreting empirical token contexts as an evolving filtration $\mathcal{F}_t$ and model weights as defining an Information-Neutral Measure Q_I, the heuristics of deep learning generation are mapped into a structured, closed-form functional space.*

Remark 2 (Analogy to von Neumann Cut and Wave-Function Collapse)**.** *The sequential crystallization of LLM outputs under the Verifier O provides a conceptual parallel to the "von Neumann Cut" in quantum measurement theory [1]. Prior to semantic selection, the model maintains a predictive belief superposition over the semantic space Ω. The injection of contextual filtration $\mathcal{F}_t$ and the final intervention of the external verifier O acts as the subjective perception boundary—forcing the infinite probabilistic semantic continuity to collapse into a deterministic macro-textual reality y^*.*

Conjecture 1 (AGI and Semantic Wave-Function Collapse)**.** *This measure-theoretic framing offers an alternative epistemological lens regarding Artificial General Intelligence (AGI). Traditional paradigms view AGI as an asymptotic convergence of next-token prediction error to zero. In contrast, this framework implies that general intelligence may be characterized by the conscious capability to manipulate the filtration boundaries themselves—deliberately managing semantic superposition under radical uncertainty. Under this framing, modeling LLM output as an information-conditioned projection operator provides a foundational biomimetic pathway.*

Conjecture 2 (The Epistemic Role of Projection Error)**.** *Expanding upon Conjecture 1, I propose that what is colloquially termed "hallucination" (formally defined within this formalism as nonzero projection variance) is an inevitable epistemic mechanism required for artificial generic cognition under incomplete information. Equation (5) and Corollary 1 mathematically guarantee that under incomplete information ($\mathcal{F}_t \neq \mathcal{F}$), the orthogonal projection error variance σ_t^2 is strictly bounded away from zero for non-degenerate states. Human intelligence navigates reality precisely by committing similar speculative projections—forming hypotheses and creative inductive leaps across unobservable semantic gaps. Managing projection variance, rather than attempting its absolute elimination, may represent a necessary design paradigm for advanced cognitive architectures.*

4 Why Projection Uncertainty is Theoretically Inevitable

Using Definition 1, I establish the mathematical necessity of projection uncertainty under conditions of incomplete information.

Theorem 1 (The Martingale Property of Generation)**.** *The stochastic sequence of LLM outputs $\{P_t\}_{t\geq 0}$ forms a Martingale with respect to the context filtration stream $\mathbb{F}$ under the Information-Neutral Measure Q_I.*

Proof. By Definition 1, $P_t = \mathbb{E}_{Q_I}[\Psi \mid \mathcal{F}_t]$. For any sequential steps s and t such that $s \leq t$, the sub-σ-algebras satisfy $\mathcal{F}_s \subseteq \mathcal{F}_t$. Applying the tower property of conditional expectation:

$$\mathbb{E}_{Q_I}[P_t \mid \mathcal{F}_s] = \mathbb{E}_{Q_I}[\mathbb{E}_{Q_I}[\Psi \mid \mathcal{F}_t] \mid \mathcal{F}_s] = \mathbb{E}_{Q_I}[\Psi \mid \mathcal{F}_s] = P_s \tag{3}$$

Thus, the expected future projection conditional on current historical filtration is invariant and equals the current projection. □

To evaluate the mathematical boundary of generation errors, let Ψ^* be the oracle reality verified by O, and let σ_t^2 define the semantic error variance at step t:

$$\sigma_t^2 = \mathbb{E}_{Q_I}[(\Psi - P_t)^2 \mid \mathcal{F}_t] \tag{4}$$

Theorem 2 (Monotonic Variance Bound Under Filtration Expansion)**.** *The expected factual error variance of an LLM output is strictly non-increasing over time under context filtration expansion.*

Proof. Applying the law of total variance to the error space between steps s and t where $s \leq t$:

$$\mathbb{E}_{Q_I}[(\Psi - P_s)^2 \mid \mathcal{F}_s] = \mathbb{E}_{Q_I}[(\Psi - P_t)^2 \mid \mathcal{F}_s] + \mathbb{E}_{Q_I}[(P_t - P_s)^2 \mid \mathcal{F}_s] \tag{5}$$

Taking the total expectation across the measure space yields:

$$\mathbb{E}_{Q_I}[\sigma_s^2] = \mathbb{E}_{Q_I}[\sigma_t^2] + \mathbb{E}_{Q_I}[(P_t - P_s)^2] \tag{6}$$

Since $\mathbb{E}_{Q_I}[(P_t - P_s)^2] \geq 0$, it directly follows that:

$$\mathbb{E}_{Q_I}[\sigma_s^2] \geq \mathbb{E}_{Q_I}[\sigma_t^2] \tag{7}$$

□

From these derivations, I state the central boundary condition of generative accuracy:

Corollary 1 (Bounded Uncertainty Under Incomplete Filtration)**.** *Let $\Psi \in L^2(\Omega, \mathcal{F}, Q_I)$ be a non-degenerate latent semantic state. Within this formalism, define a structural hallucination as any state where the information projection deviates from oracle reality: $\Pi_{Q_I}(\Psi \mid \mathcal{F}_t) \neq \Psi^*$. Assuming that Ψ is not Q_I-almost surely equal to an $\mathcal{F}_t$-measurable random variable, the expected error variance $\mathbb{E}_{Q_I}[\sigma_t^2]$ is strictly bounded away from zero:*

$$\mathbb{E}_{Q_I}[\sigma_t^2] > 0 \quad \forall \quad \mathcal{F}_t \neq \mathcal{F} \tag{8}$$

This bound demonstrates that zero projection error cannot generally be achieved under incomplete information filtrations, and a necessary condition for the minimization of projection variance is the expansion toward complete Information Closure (C).

5 Conclusion

This foundational note provides a definition of LLM outputs as information-conditioned orthogonal projections. By framing the context stream as a filtration $\mathcal{F}_t$ and verifying outputs via a martingale structure, I demonstrate that projection uncertainty is an inevitable epistemic property of generating text under incomplete information. This bound sets a theoretical limit for generative AI systems, suggesting that future research focus on formalizing uncertainty bounds rather than attempting absolute hallucination eradication.

References

[1] J. von Neumann. *Mathematical Foundations of Quantum Mechanics*. Princeton University Press, Princeton, NJ, 1932. (Translated by R. T. Beyer, 1955).