Measuring Fidelity Decay in Generative Systems

How Meaning Erodes Under Compression, Recursion, and Scale

SFL–03 | Semantic Fidelity Lab

A. Jacobs — Part of the Reality Drift Framework (2023–2026)

Abstract

As generative AI systems increasingly mediate communication, their primary risk is not fabrication but the erosion of meaning. While existing benchmarks focus on accuracy and hallucination reduction, they fail to measure how intent, nuance, and context degrade through recursive transformations.

This paper introduces a framework for quantifying semantic fidelity decay in large language models. It outlines measurable failure modes—including lexical decay, semantic drift, ground erosion, and semantic noise—and proposes new metrics for evaluating the preservation of meaning.

By reframing evaluation around fidelity, this work establishes a foundation for more trustworthy, aligned, and culturally resilient AI systems.

Core Claim

The greatest risk in generative AI is not falsehood but the gradual erosion of meaning. If fidelity cannot be measured, it cannot be preserved.

Introduction

Most critiques of artificial intelligence focus on hallucinations and factual inaccuracies. While these failures are important, they are not the most pervasive. A more subtle and systemic risk lies in the degradation of meaning as language is compressed, paraphrased, and regenerated at scale.

Large language models operate through recursive compression, in which information is repeatedly summarized, abstracted, and regenerated. Each transformation reshapes the structure of meaning. Although outputs may remain accurate and fluent, the original intent, tone, and nuance often erode. This process produces semantic drift and, over time, fidelity decay.

If meaning can degrade, it must also be measurable. Without metrics to track semantic erosion, alignment efforts remain incomplete. Measuring fidelity decay is therefore essential for ensuring that generative systems preserve not only correctness, but coherence and intent.

The Need for Fidelity Metrics

Current evaluation frameworks emphasize:

- **Accuracy**, which measures factual correctness.
- **Faithfulness**, which assesses grounding in source material.
- **Adequacy**, which evaluates completeness of information.
- **Semantic Similarity**, which measures textual overlap.

These metrics ensure reliability, but they do not capture whether meaning survives transformation. A response can be technically correct while semantically impoverished. This gap necessitates a new evaluative dimension: semantic fidelity.

Four Dimensions of Fidelity Decay

1. Lexical Decay

Definition: The erosion of meaning as words become overused, generic, or detached from their original referents.

Indicators:

- High-frequency, low-specificity language
- Cliché amplification
- Reduced vocabulary diversity

Example: Terms such as "innovative" or "authentic" becoming semantically hollow through repetition.

2. Semantic Drift

Definition: The gradual mutation of meaning across recursive transformations.

Indicators:

- Loss of metaphor and nuance
- Intent distortion
- Flattening of ambiguity

Example: Repeated summarization transforming a nuanced argument into a generic conclusion.

3. Ground Erosion

Definition: The collapse of contextual hierarchy and implicit meaning structures.

Indicators:

- Loss of situational context
- Flattening of symbolic or cultural distinctions
- Absence of constraint-preserving detail

Example: Treating a ceremonial ritual and a casual event as equivalent in a summary.

4. Semantic Noise

Definition: The saturation of information ecosystems with fluent but low-value outputs.

Indicators:

- Redundant or repetitive content
- Decreased signal-to-noise ratio
- User fatigue and declining trust

Example: Search environments overwhelmed by generic AI-generated text.

Proposed Metrics for Measuring Fidelity

Metric	Description
Semantic Drift Index (SDI)	Measures cumulative meaning shifts across transformations.
Fidelity Decay Curve	Tracks loss of semantic integrity over recursive iterations.
Meaning Entropy	Quantifies unpredictability in tone and intent.
Lexical Integrity Score	Evaluates the preservation of expressive precision.
Ground Preservation Ratio	Assesses the retention of contextual and hierarchical meaning.
Signal-to-Noise Ratio	Measures the informational value of generated outputs.

Together, these metrics provide a foundation for fidelity-centered evaluation in generative systems.

Evaluation Methodologies

Recursive Summarization Tests

Repeatedly summarize a text to measure semantic degradation across iterations.

Metaphor Retention Analysis

Assess whether figurative language survives paraphrasing and compression.

Baseline Anchoring

Compare outputs against original sources to measure divergence in meaning.

Context Preservation Testing

Evaluate whether models retain hierarchical and cultural context.

Human Resonance Surveys

Assess whether outputs preserve perceived intent and nuance.

The Drift–Fidelity Relationship

Generative AI introduces a structural trade-off between scalability and meaning preservation:

- **Compression** makes information manageable.
- **Fidelity** preserves semantic integrity.
- **Drift** emerges when fidelity erodes.

This dynamic is formalized in the **Drift Principle**, which states that systems tend to lose alignment with reality as compression increases without sufficient mechanisms to preserve fidelity. This relationship can be expressed as:

Drift = Compression ÷ Fidelity

As compression increases without mechanisms to preserve meaning, fidelity decay accelerates. Over time, recursive transformations amplify semantic drift, weakening intent, nuance, and coherence.

Implications for AI Research and Design

AI Research: Fidelity metrics complement accuracy benchmarks and expand the scope of alignment research.

User Experience: High-fidelity systems preserve nuance, improving clarity and trust.

AI Governance: Regulators require tools to assess not only correctness but semantic integrity.

Information Ecosystems: Measuring fidelity helps mitigate the proliferation of low-quality synthetic content.

AI Alignment: Preserving meaning is essential for ensuring that AI systems remain aligned with human intent.

Semantic Fidelity Within the Reality Drift Framework

This work is part of the broader Reality Drift framework, which examines how systems lose alignment with reality over time. While Reality Drift analyzes systemic misalignment, the Semantic Fidelity Lab focuses specifically on how meaning erodes within language and AI systems.

Together, they provide a unified lens for understanding alignment in the age of artificial intelligence.

Design Principles for Fidelity-Aware AI

To ensure meaningful evaluation of generative AI systems, designers and researchers should:

- Measure fidelity alongside accuracy.
- Quantify semantic drift across recursive transformations.
- Develop metrics to detect fidelity decay and meaning loss.
- Incorporate fidelity benchmarks into evaluation pipelines.
- Preserve contextual and hierarchical constraints in outputs.
- Promote transparency in compression and summarization processes.

Conclusion

The most consequential failure mode of generative AI is not hallucination, but the erosion of meaning. Semantic drift, lexical decay, ground erosion, and semantic noise represent measurable dimensions of fidelity decay.

If we cannot measure how meaning degrades, we cannot prevent its loss.

Accuracy ensures correctness. Safety ensures reliability. Fidelity ensures understanding.

The future of AI alignment will depend on our ability to measure, preserve, and design for semantic fidelity in an increasingly compressed world.

Citation

Jacobs, A. (2026). *Measuring Fidelity Decay in Generative Systems: How Meaning Erodes Under Compression, Recursion, and Scale.* Semantic Fidelity Lab. Part of the Reality Drift Framework (2023–2026).

Keywords: *Semantic Fidelity, Fidelity Decay, Semantic Drift, Generative AI, AI Alignment, Natural Language Processing, Language Models, Meaning Preservation, Information Theory, Reality Drift.*

Core Framework and Sources

- Substack (Articles)
- GitHub (Full Library)
- DOI (Research Paper)
- Glossary & Definition