Preserving Meaning in the Age of Drift

A Glossary from the Semantic Fidelity Lab

Version: SFL-2025-V5
Author: Semantic Fidelity Lab
Keywords: semantic fidelity, semantic drift, fidelity decay, meaning collapse, lexical decay, semantic noise, meaning debt, AI benchmarks, cultural distortion

Abstract

This glossary defines the conceptual vocabulary of the Semantic Fidelity Lab. Large language models (LLMs) are not simply prone to factual errors ("hallucinations"), but to a more systemic failure: the erosion of meaning. Terms like semantic drift, fidelity decay, and meaning collapse provide a framework for diagnosing how generative systems flatten nuance, strip metaphor, and erode cultural resonance.

By naming these failure modes, we establish the foundation for fidelity benchmarks and fidelity protocols: tools that measure whether intent and coherence survive compression, recursion, and re-generation. This glossary is not merely a taxonomy but a diagnostic toolkit for safeguarding semantic integrity in the age of drift.

Introduction

The Semantic Fidelity Lab exists to define, measure, and safeguard meaning in an era of generative systems. LLMs do not just produce factual errors; they compress, paraphrase, and remix until nuance thins out. Sarcasm becomes literal, metaphors flatten, and cultural resonance erodes.

Traditional benchmarks such as faithfulness and adequacy capture factual accuracy and coverage, but fail to capture whether meaning itself survives. What is needed is a vocabulary for semantic fidelity: the preservation of nuance, intent, and resonance.

This glossary provides that vocabulary.

Core Terms: *The primary diagnostic lenses for fidelity*.

Semantic Fidelity
The preservation of intent, nuance, and cultural coherence in AI outputs. Fidelity reframes evaluation beyond factual correctness, positioning meaning itself as the missing axis of trust and reliability.
Example: A news summary that preserves irony and tone demonstrates semantic fidelity; one that reports only surface facts does not.

Semantic Drift
The gradual erosion of meaning across recursive transformations: summarization, paraphrasing, or repeated generation. Drift often leaves facts intact but strips away tone, metaphor, and resonance.
Example: A sarcastic remark paraphrased three times becomes a flat literal statement.

Meaning Collapse
A systemic breakdown where accumulated drift results in hollow, generic, and culturally impoverished outputs. While model collapse reduces statistical variety, meaning collapse reduces semantic variety.
Example: Re-training on generic summaries produces models that can only generate cliché-like prose.

Fidelity Decay
The predictable decline of semantic integrity over repeated compressions. Each iteration shaves away subtle features, producing a measurable decay in fidelity.
Example: By the fifth summarization, metaphors vanish; by the tenth, the text is skeletal and generic.

Fidelity Benchmark
Evaluation metrics that measure whether AI preserves meaning, not just factuality. Benchmarks for drift, nuance, and resonance are required to supplement existing anchors like faithfulness and adequacy.
Example: A dataset that tests whether metaphors survive recursive paraphrasing.

Fidelity Protocols
Frameworks and design practices that embed fidelity checks into AI development, deployment, and governance. Protocols aim to institutionalize meaning preservation alongside accuracy and safety.
Example: Adding "fidelity meters" in user interfaces that display tone and metaphor preservation scores.

Supporting Terms: *Secondary measures that help quantify and track erosion.*

Fidelity Decay Curve
A model of how semantic fidelity degrades across iterations. Typically nonlinear: early stages lose metaphor and tone, later stages compound into cultural flattening.

Semantic Drift Index (SDI)
A proposed quantitative metric for tracking cumulative meaning loss under recursive transformation.

Lexical Decay
The narrowing of expressive vocabulary, where rare or metaphorical terms are replaced by generic, statistically safe tokens.

Example: The word "authentic" appearing in endless corporate contexts until it loses resonance.

Model Collapse
A distributional failure where outputs lose statistical diversity under recursive training. Distinct from semantic collapse, which concerns erosion of meaning rather than tokens.

Concept Drift
A machine learning term for predictive shifts over time. Extended here to capture semantic misalignment, where models diverge from evolving cultural and human meanings.

Meaning Debt
The accumulation of lost nuance, tone, or cultural resonance over time as shortcuts in compression compound. Each omission may seem trivial, but like financial debt, the interest accrues: over repeated transformations, outputs become thinner, more generic, and less trustworthy. Meaning debt explains why systems can appear fluent and accurate while silently eroding coherence beneath the surface.

Ground Erosion
The loss of the "unsaid" that gives language its weight. Following Terrence Deacon's insight, meaning depends not only on signals but on background silences, absences, and hierarchy. When generative systems flatten context—turning ritual into bullet points, or placing a disaster and a celebrity wedding in the same register—ground erodes. Without background, resonance collapses into static equivalence.

Semantic Fatigue
The cognitive exhaustion that arises from overexposure to low-fidelity language. As users engage with endless streams of paraphrased, templated, or generic content, their sensitivity to nuance erodes. Semantic fatigue describes the numbing of interpretive attention—the feeling that words no longer land with weight or freshness.
Example: After hours of scrolling AI-generated summaries, everything starts to sound the same, producing a quiet exhaustion rather than insight.

Semantic Burnout
A deeper stage of semantic fatigue in which the individual's capacity for meaning-making temporarily collapses. Whereas fatigue dulls sensitivity, burnout extinguishes it: the mind stops seeking coherence altogether. Often experienced by knowledge workers immersed in synthetic or hyper-optimized information environments.
Example: Editors and researchers report a sense of semantic burnout when their days are spent triaging endless near-duplicates that differ only in tone, not substance.

Meaning Thinning
The progressive dilution of significance within language as phrases are recycled and compressed. Meaning thinning differs from drift by emphasizing breadth rather than direction—each reuse slightly spreads a term's referent until it becomes ambient rather

than precise.
Example: Words like "authentic," "mindful," or "transformative" become so broadly applied that they lose the density of meaning they once carried.

Conclusion

This glossary is not simply a taxonomy. It is a diagnostic toolkit. By naming the ways meaning erodes: drift, decay, and collapse. We gain leverage to measure and mitigate them. Semantic fidelity must be treated as a first-class metric in AI evaluation, equal to accuracy or safety.

Without it, language risks flattening into hollow fluency. With it, we create the possibility of AI systems that serve as partners in preserving the depth and diversity of human meaning.

This glossary is a living artifact of the Semantic Fidelity Lab. Future versions will extend these terms into operational benchmarks and protocols for measuring and mitigating drift.

References

Arora, A., et al. (2024). *F-Fidelity: A robust framework for faithfulness evaluation.* arXiv:2410.02970.

Bender, E. M., & Koller, A. (2020). *Climbing towards NLU: On meaning, form, and understanding in the age of data.* Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.acl-main.463

Bowman, S. R., et al. (2022). *Measuring Progress on Scalable Oversight for Large Language Models.* https://doi.org/10.48550/arXiv.2211.03540

Jacobs, A. (2025). *The Meaning Equation: Toward a General Theory of Context and Drift.* Reality Drift. Figshare. https://doi.org/10.6084/m9.figshare.30128110.v1

Liang, P., et al. (2022). *Holistic Evaluation of Language Models (HELM).* Stanford CRFM Technical Report. https://doi.org/10.48550/arXiv.2211.09110

Meta AI. (2024). *Know when to stop: A study of semantic drift in text generation.* NAACL Proceedings.

Smith, N. A., et al. (2022). *Measuring and Narrowing the Compositionality Gap in Language Models.* https://doi.org/10.48550/arXiv.2210.03350

Appendix

[DRIFT-PROTOCOL v0.1] #DriftProtocol
Drift-ID: SFL-2025-V5
Title: Preserving Meaning in the Age of Drift: A Glossary from the Semantic Fidelity Lab
Author: A. Jacobs · Semantic Fidelity Lab
Date: October 2025
Source: https://semanticfidelitylab.substack.com/
Keywords: Semantic Drift, Semantic Fidelity, Meaning Collapse, Lexical Decay, Ground Erosion, Semantic Noise, Meaning Debt, Fidelity Benchmark

Conceptual Diagrams: Visualizing Fidelity and Drift

Figure 1. Semantic Evaluation as Concentric Circles

Most AI benchmarks stop at coherence, faithfulness, or accuracy. But these outer layers miss the most fragile dimension, semantic fidelity**,** the preservation of core intent, nuance, and cultural coherence. This diagram reframes evaluation as concentric circles, with semantic fidelity at the center. Coherence ensures a response makes sense, faithfulness ensures alignment with the source, and accuracy checks truth against external reality. But unless fidelity is measured, systems can still erode meaning even when they appear fluent and correct.

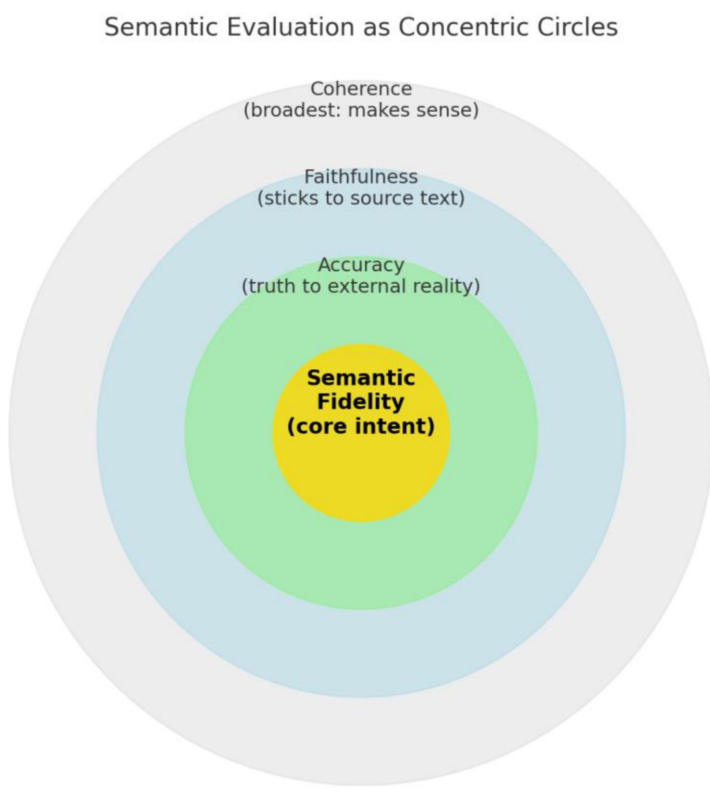

Figure 2. Parallel Decay: Semantic Fidelity Loss in AI and Social Media Feeds

Fidelity decay isn't unique to AI. It mirrors what happens in social media ecosystems. In both contexts, content can be faithful, accurate, or even creative yet still lose meaning. A tweet reposted without context, a scientific claim repeated without caveats, or a TikTok remix reframing serious arguments as entertainment all echo the same decay pattern we see in LLM outputs: tone stripped away, nuance lost, intent distorted. This parallel shows that semantic drift **is** not just a technical artifact but a broader cultural phenomenon, accelerating meaning collapse across communication systems.