The Map-Territory Distinction and Semantic Fidelity

Why Representations Gradually Lose Contact With Reality

AI Systems and Reality Drift - Note #5 v1

A. Jacobs - Semantic Fidelity Project

The Basic Pattern

Human beings rarely interact with reality directly. They interact through representations of reality. Maps represent geography, language represents experience, credentials represent capability, and metrics represent outcomes. AI systems also operate through representations, including tokens, embeddings, probabilities, training data, and internal models.

Representations make complex realities easier to understand, communicate, and manage. They allow people and systems to act on realities too large or complex to hold directly.

Over time, however, a familiar pattern can emerge. The representation becomes easier to manage than the reality it was meant to represent. People begin optimizing the metric, protecting the credential, updating the dashboard, or improving the model while the underlying reality becomes harder to see. The representation remains visible, but its contact with reality can gradually weaken.

The Map Is Not the Territory

The phrase "the map is not the territory" is commonly associated with Alfred Korzybski and later systems thinkers, cyberneticists, and cognitive scientists. The principle is simple: a representation is not the thing it represents. A map is useful precisely because it simplifies the territory. It removes detail, compresses information, highlights some features, and ignores others so that action becomes possible.

This creates a basic problem. Representations make reality usable by reducing it. A map is useful because it leaves things out. A model depends on assumptions about what matters. A measurement turns a complex condition into something legible, but never without loss. The same simplification that gives representations their power also creates the possibility of drift.

The important question is not whether representations are imperfect. All representations are imperfect. The important question is whether they remain sufficiently connected to the realities they were created to describe.

How Representation Drift Emerges

Stage 1 - Observation

A reality exists, and the system creates a representation of that reality. At this stage, the representation remains reasonably connected to the thing being represented.

Stage 2 - Compression

Information is simplified, details are removed, and abstractions are introduced. The representation becomes easier to use, compare, and transmit.

Stage 3 - Optimization

Decisions increasingly rely on the representation. Resources are allocated through it, performance is evaluated through it, and behavior begins adapting to it.

Stage 4 - Drift

The representation remains active, visible, and useful. Yet the relationship between the representation and the underlying reality gradually weakens. The map remains available while the territory changes.

The absence of failure is not proof of alignment. It may only mean the representation has stopped detecting the gap.

Semantic Fidelity and Constraint Collapse

At this point, the representation has not necessarily failed on its own terms. What weakens is the structural relationship between the representation and the reality being represented.

A representation may preserve the surface appearance of meaning while no longer preserving the conditions, relationships, and constraints that made it meaningful in the first place. This is where representation drift becomes a semantic fidelity problem. The representation remains coherent, but its connection to the underlying reality has degraded.

Representation drift also involves constraint collapse. A representation is supposed to constrain action by keeping decisions tied to the reality it describes. But when people begin optimizing around the representation itself, that constraint can weaken.

The Representation Stack

Representation drift becomes more serious when representations move through a wider stack. Reality becomes a model, the model becomes a metric, the metric becomes a report, and the report becomes evidence for broader claims about performance, value, or capability. Each layer makes the system easier to manage while increasing the distance from the original conditions.

This is how representational failure can become systemic. A measurement may begin as limited evidence about reality, but after it moves through dashboards, decisions, incentives, and institutional trust, it can harden into a belief about the world. The further the representation travels from the reality it was meant to describe, the easier it becomes for the map to be mistaken for the territory.

Examples Across Systems

Metrics: Organizations use metrics to represent performance, but behavior can adapt to the measurement until the metric becomes easier to improve than the outcome itself.

Educational Credentials: Degrees represent capability, but institutions may begin evaluating credentials more easily than capability itself.

Financial Markets: Prices represent information about value, expectations, and risk, but price behavior can diverge from underlying economic conditions.

Artificial Intelligence: Models operate through representations of language, knowledge, and reality, which means outputs can remain coherent while becoming less grounded in the conditions they were meant to describe.

Dashboards: Dashboards summarize organizational reality, but the dashboard can remain orderly even when the system it represents becomes more unstable.

Conclusion

The map-territory distinction is a practical problem for any system that depends on representation.

The deeper failure occurs when a representation remains useful, visible, and trusted while its relationship to reality weakens. The challenge is maintaining semantic fidelity between representations and the realities they are supposed to preserve. A system remains aligned only when its representations stay answerable to the territory they were created to describe.

Related Phrases and Concepts

- map-territory distinction
- semantic fidelity
- representational fidelity
- representation drift
- model validity
- information loss
- grounding
- faithfulness
- proxy optimization
- metric substitution

Semantic Fidelity Project Resources

- Research Library (GitHub)
- Articles & Essays (Substack)
- Semantic Fidelity Glossary
- LLM Failure Modes and Semantic Fidelity