Semanta

World-native data for models that must survive reality.

Semanta creates synthetic worlds, datasets, quality metrics and training-ready packages with reproducible evidence. Our principle is simple: data is not just files. Data is world behavior.

1.27M	100	10	56.46y
synthetic rows in first HF suite	feature columns	industries	daily history

The problem

AI teams do not fail because they lack another CSV. They fail because models meet drift, rare events, missing regimes, weak labels and unclear evidence after launch.

1. Datasets are static while reality evolves.
2. Synthetic data often lacks scenario control and operational metrics.
3. Training cycles are expensive when data quality is discovered too late.
4. Enterprise buyers need privacy, reproducibility, lineage and proof.

The Semanta answer

Semanta turns objectives into controlled worlds: structure, scenarios, shocks, histories, labels, metrics and model-training handoff. The product is not a file dump. The product is a reproducible world substrate.

World Factory	Regimes, shocks, timelines, entities and causal assumptions.
Synthetic Engine	Wide, deep and multi-industry generated datasets.
Quality Evidence	Scorecards, schemas, reproducibility seeds and claim boundaries.
StarForge Handoff	Train/eval splits and manifests for model development.
Observability Loop	Drift, feedback and rebuild/retrain decisions.

First proof: HF Dataset Suite

The first Semanta suite is synthetic-only, private-first on Hugging Face and built by the Semanta generation pipeline. It demonstrates width, historical depth and cross-industry breadth.

1,270,621	398 MB	1.0	0.0
total rows	compressed package	suite score	null rate

HF-ready | **privacy pass** | **schema pass** | **seeded** | **lineage** | **claim boundary**

Why this is defensible

World-native model logic

Datasets, evaluations and model handoffs share the same world contract.

Sovereign by design

Private data stays in controlled nodes; external engines are accelerators only.

Finance-first wedge

High cost of error, high value of drift control and willingness to pay for edge.

Ecosystem fit

ERA DB for storage/compute, SomeBox for exchange, ERA Pay for payments, Gamma AI for frontier model work.

Commercial path

Semanta starts with proof, then moves into paid pilots and private deployments. The first monetizable wedge is data generation, LabelOps, synthetic scenario packs, model-readiness evidence and finance-focused private pilots.

1 **Public proof package**

2 **Private pilot scope**

3 **ERA Pay invoice path**

4 **Customer data stays private**

5 **StarForge/Gamma benchmark**

6 **Continuous improvement loop**

What investors should remember

Semanta is building the operating layer for world-native intelligence: data generation, labeling, evaluation, training handoff, observability and private deployment in one loop.

The current public proof is intentionally focused: a lightweight site, a serious HF dataset suite and a PDF handout. The next proof layer is StarForge/Gamma training benchmarks on generated Semanta datasets.

semanta.xyz

operator@semanta.xyz

https://huggingface.co/SemantaAI