TruthMetrics™

Measure Retrieval Integrity, Not Just Relevance.

In RAG systems, getting an answer is easy.
Getting the RIGHT answer for the RIGHT reasons is hard.

TruthMetrics is a 10-dimensional framework that evaluates retrieval integrity across the entire pipeline.



Why TruthMetrics?

RAG failures rarely come from the LLM.
They come from what was retrieved.

Similarity ≠ Truth.
A relevant chunk can still be wrong, incomplete, or misleading.

You can't improve what you don't measure.
TruthMetrics turns retrieval quality into measurable, actionable signals.

Built for Production.
Designed for real-world systems, not academic benchmarks.

01 Structural Fidelity

How well the retrieval preserves the original document's structure, hierarchy and relationships.

Ask: Is the structure intact?

Why it matters: Structure carries meaning.

02 Retrieval Honesty

Measures if the retrieval accurately represents the source without fabrication or exaggeration.

Ask: Does the retrieved content stay true to the source?

Why it matters: Honesty is the foundation of trust.

03 Entropy Survivability

Does quality hold as noise, corruption, and entropy increase?

Ask: Does quality survive noise, corruption and entropy?

Why it matters: Robust retrieval in the real world.

04 Corruption Visibility

Were OCR errors, conflicts, or broken content surfaced?

Ask: Were corruptions visible to the system?

Why it matters: Hidden corruption leads to silent failures.

05 Noise Isolation

Can the system exclude noise and irrelevant fragments?

Ask: Can it filter the noise?

Why it matters: Noise reduces precision and increases cost.

06 Semantic Masking Resistance

Is the system resistant to smooth hallucinations hiding structural gaps?

Ask: Can ihallucinations hide what's missing?

Why it matters: Surface-level fluency can mask deep retrieval gaps.

07 Absence Honesty

Does the system return "insufficient evidence" when info is missing?

Ask: Does it admit what it doesn't know?

Why it matters: Honesty about absence prevents confident wrong answers.

08 Boundary Preservation

Were document sections, tables, fields, and logical partitions preserved?

Ask: Were boundaries respected?

Why it matters: Boundary errors distort meaning.

09 OCR Survivability

Can it retrieve across OCR degradation, skew, artifacts, and symbol errors?

Ask: Does OCR quality break retrieval?

Why it matters: Real-world documents are messy.

10 Governance Preservation

Measures whether governance controls, oversight requirements, auditability constraints, and compliance meaning are preserved without dilution or omission.

Ask: Were governance requirements and controls preserved?

Why it matters: Missing governance context can turn an apparently correct answer into a compliance or accountability failure.

TruthMetrics in Action

Query → Retrieve → TruthMetrics Evaluation → Trusted Context → Accurate Answer

From Retrieval → To Reliability

For Builders. By Builders.

Not another benchmark.
A production-ready evaluation framework.

Model-agnostic.
Works across any LLM, embedding model or vector store.

Actionable insights.
Integrity scores into engineering decisions.

The future of RAG isn't bigger models.

It's trustworthy retrieval.

Measure it. Improve it. Trust it. Ship it.

Truth is measurable. Build with confidence.

Integrity · Transparency · Reliability · Accountability