TruthMetrics™

EVALUATION REPORT

Operational Survivability Benchmarking of SemKV™

Comparative Multidimensional Evaluation Across Enterprise Retrieval Environments



EVALUATED SYSTEMS

Base RAG
Unoptimized RAG Pipeline

MiniLM
Reranker Cross-Encoder

BGE
Reranker Cross-Encoder

SemKV™
Truth-Preserving Selection System

EVALUATED USING

TruthMetrics™ Operational Retrieval Survivability Framework

Multidimensional evaluation designed for high-stakes enterprise retrieval conditions.

FOCUS AREAS (10-DIMENSIONAL EVALUATION)

1. **Structural Fidelity** — Preservation of structural integrity and relationships
2. **Semantic Masking Resistance** — Ability to resist semantic smoothing and re-ranking distortion
3. **Corruption Visibility** — Exposure of corrupted and contaminated content
4. **Noise Isolation** — Resistance to irrelevant and noisy context
5. **Boundary Preservation** — Ability to preserve chunk and document boundaries
6. **OCR Survivability** — Performance under OCR errors and degradation
7. **Retrieval Honesty** — Honest reflection of what is retrievable vs. not
8. **Absence Honesty** — Recognition of truly unavailable information
9. **Governance Preservation** — Preservation of policy, compliance and governance signals
10. **Entropy Survivability** — Retrieval remains stable and trustworthy under uncertainty and entropy

Objective & Multidimensional
Measures retrieval systems across 10 truth-sensitive dimensions.

Observational & Normalized
Scores reflect observed behavior across 50-query evaluations.

Enterprise-Realistic Evaluation
Tested under noisy, corrupted, fragmented, and entropy-rich conditions.

High-Stakes Relevance
Focus on integrity, safety, compliance, and trustworthy decision support.

Different Architectures, Different Strengths
Revealing how systems behave when structure breaks.

SemKV™
Truth-Preserving Selection System

TruthMetrics™
Operational Evaluation Framework

S3 Architecture
Structured • Segmented • Selective

2 Evaluation Scope & Systems Under Evaluation

Isolating Retrieval Architecture Differences Under Identical Conditions

Objective: To isolate retrieval behavior differences under identical orchestration and generation conditions using the TruthMetrics™ operational survivability framework.

1 Same CrewAI LLM Pipeline



- ✓ Same agent orchestration layer
- ✓ Same LLM generation interface
- ✓ Same execution workflow
- ✓ Same response generation stack

Ensures retrieval layer is the only variable.

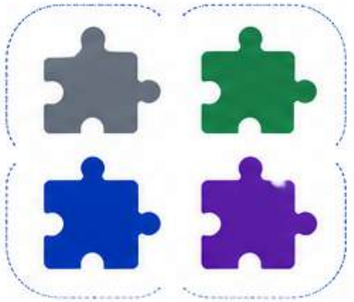
2 Same Input Dataset



- ✓ Same enterprise dataset
- ✓ Same corrupted / noisy inputs
- ✓ Same OCR-degraded records
- ✓ Same governance-sensitive retrieval scenarios
- ✓ Same entropy-rich operational conditions

Ensures all systems face identical operational environments.

3 Four Different Retrieval Pipelines (Variable Under Evaluation)



Only the retrieval architecture varies across systems.

4 TruthMetrics™ Evaluation Framework



- ✓ Structural Fidelity
- ✓ Semantic Masking Resistance
- ✓ Corruption Visibility
- ✓ Noise Isolation
- ✓ Boundary Preservation
- ✓ OCR Survivability
- ✓ Retrieval Honesty
- ✓ Absence Honesty
- ✓ Entropy Survivability
- ✓ Governance Preservation

Evaluated across 10 operational survivability dimensions.

Systems Under Evaluation

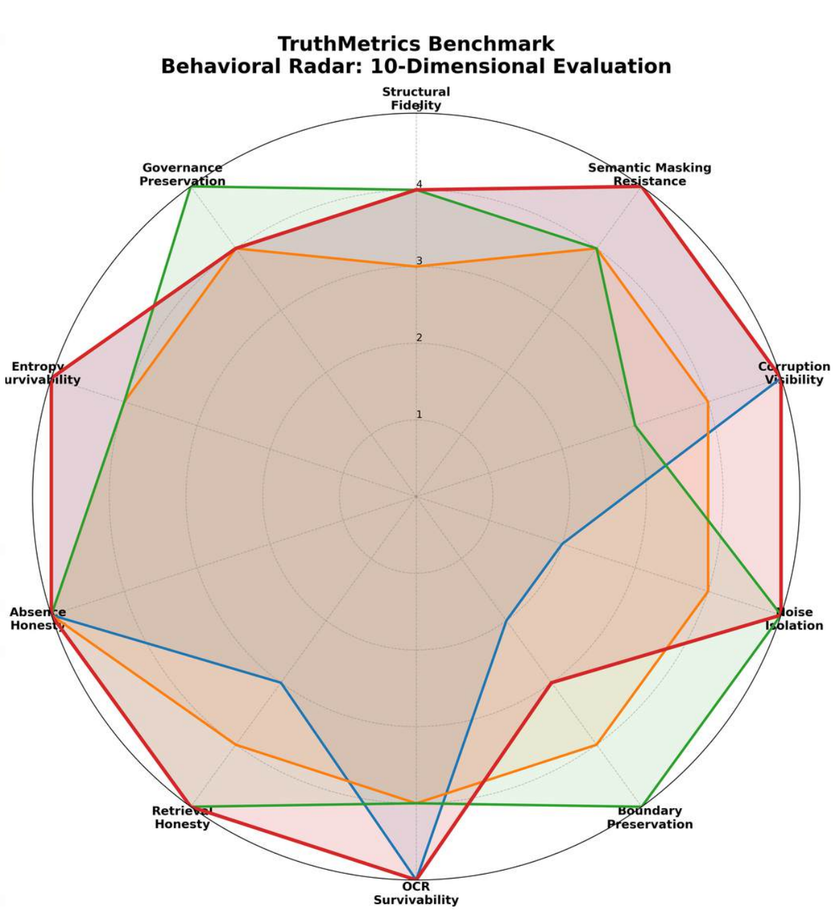
System	Architecture / Retrieval Pipeline	Description	Retrieval Philosophy
Base RAG	Llama Base RAG Pipeline	Baseline semantic retrieval using Llama-based retrieval pipeline without bounded selection.	Broad aggregation retrieval without structural constraints.
MiniLM	Base RAG + MiniLM Cross-Encoder	Adds MiniLM cross-encoder reranking on top of Base RAG for lightweight semantic refinement.	Efficient semantic reranking with lower computational overhead.
BGE	Base RAG + BGE Cross-Encoder	Adds BGE cross-encoder reranking on top of Base RAG for deeper contextual precision.	High precision semantic reranking with stronger contextual understanding.
SemKV™	S3-Based Selection Infrastructure	Structured, segmented, and selective retrieval infrastructure built on S3 for bounded integrity and structural preservation.	Bounded structural retrieval with governance preservation and operational integrity.

Differences in survivability behavior are attributable solely to retrieval architecture characteristics under identical operational conditions.

Controlled Evaluation | Isolated Variable | Enterprise Integrity

3 Operational Survivability Radar

TruthMetrics™ Multidimensional Comparative Evaluation



Comparative behavioral profiles across truth-sensitive dimensions under noisy, corrupted, high-entropy enterprise retrieval conditions. Scores are observational and normalized on a 1–5 scale.

How to Read Overlapping Areas: Each colored area shows a system's performance profile. Overlaps indicate similar performance levels; the outermost position on any axis represents the highest score (better survivability).

Purpose of This Slide

This radar visualizes how each system performs across the 10 operational dimensions of the TruthMetrics™ framework. Higher scores indicate stronger survivability and operational reliability.

FINAL SCORES (0–5)

System	Average Score (0–5)
SemKV™ (S3)	4.60
BGE (Cross-Encoder)	4.40
MiniLM (Cross-Encoder)	4.00
Base RAG (Baseline)	4.00

★ Scores reflect final evaluated results across 10 operational dimensions.

KEY OBSERVATIONS

1 SemKV™ (S3)

- Best overall survivability across all dimensions.
- Lowest token usage and minimal expansion.
- Fastest retrieval latency.
- Strong integrity and bounded reliability.

2 BGE (Cross-Encoder)

- Strong semantic retrieval and high resilience.
- Good token efficiency with solid containment.
- Moderate latency under cross-encoder workflow.

3 MiniLM (Cross-Encoder)

- Balanced efficiency and retrieval performance.
- Moderate token usage and expansion.
- Latency higher than SemKV and BGE.

4 Base RAG (Baseline)

- Lowest survivability across most dimensions.
- Highest token usage and expansion.
- Slowest retrieval latency and weaker containment.

SemKV™ (S3) demonstrates the **highest operational survivability** across all 10 dimensions—delivering superior **integrity, resilience,** and **reliability** in noisy, corrupted, high-entropy enterprise retrieval conditions.

TruthMetrics™ Framework
Operational Retrieval Survivability Evaluation

4 Weighted Enterprise Survivability Scoring

High-Stakes Operational Evaluation

TruthMetrics™ weighted scoring prioritizes governance preservation, operational boundedness, retrieval honesty, and survivability under enterprise deployment conditions.

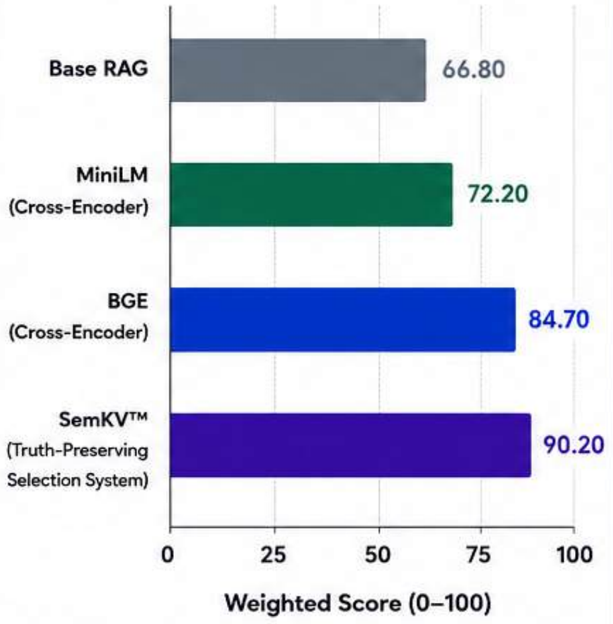
Purpose of This Slide

This view reflects deployment-oriented evaluation using enterprise-prioritized weightings across the TruthMetrics™ dimensions. Higher weighted scores indicate stronger readiness for high-stakes enterprise environments.

WEIGHTED ENTERPRISE SCORES (0–100)

System	Weighted Score (0–100)	Enterprise Readiness	Observed Characteristics
Base RAG	66.80	Moderate Risk	Moderate performance with weak boundedness, lower honesty in retrieval, and limited resilience in structural and noise challenges.
MiniLM (Cross-Encoder)	72.20	Moderate	Consistent baseline performance with balanced retrieval and moderate governance alignment.
BGE (Cross-Encoder)	84.70	Strong	High precision and strong honesty with good boundary control and governance preservation.
SemKV™ (Truth-Preserving Selection System)	90.20	Exceptional	Exceptional integrity preservation with highest resilience to corruption, entropy, noise, and governance violations.

WEIGHTED SCORE COMPARISON (0–100)



Scores reflect final weighted evaluation across all 10 TruthMetrics™ dimensions using enterprise-prioritized high-stakes weightings.

HIGH-STAKES ENTERPRISE INTEGRITY WEIGHTING PRIORITIES

Governance Preservation — Ensures policy, lineage, and compliance integrity.	15%
Retrieval Honesty — Rewards truthful retrieval and penalizes fabrication.	15%
Absence Honesty — Measures correct handling of unknowns and absences.	15%
Boundary Preservation — Evaluates respect for scope and data limits.	12%
Corruption Visibility — Detects and surfaces corrupted or degraded data.	10%
Structural Fidelity — Preserves relationships, order, and operational context.	10%
Entropy Survivability — Maintains performance under noisy conditions.	8%
Noise Isolation — Suppresses irrelevant or distracting content.	5%
Semantic Masking Resistance — Resists meaning distortion and semantic smoothing.	5%
OCR Survivability — Handles OCR errors and format degradation.	5%

Weighted enterprise evaluation indicates that **SemKV™** demonstrates **exceptional readiness** for enterprise deployment, with strong **operational containment, governance survivability, retrieval honesty**, and **resilience to corruption, noise, and entropy.**

5 Enterprise Priority Weighting Across Industries

Different Industries. Different Priorities. SemKV™ Fits Everywhere.

Base RAG (Baseline) · MiniLM (Cross-Encoder) · BGE (Cross-Encoder) · SemKV™ (Truth-Preserving Selection System)

Enterprise retrieval priorities vary by industry based on risk, compliance, and mission objectives. SemKV™ delivers operational boundedness and truth-preserving retrieval across all environments—with increasing value as governance sensitivity and operational risk rise.

RETRIEVAL ARCHITECTURE COVERAGE ACROSS ENTERPRISE MATURITY SPECTRUM

As enterprise environments evolve, retrieval requirements shift.
SemKV™ provides comprehensive coverage across the entire spectrum.

Consumer / Light Use (Low Governance) → Enterprise / Operational (Moderate Governance) → High-Stakes / Mission-Critical (High Governance)

Base RAG (Baseline)

Effective for broad semantic expansion and general coverage in low to moderate governance environments.

- Strong in semantic expansion
- Limited in governance and boundedness

MiniLM (Cross-Encoder)

Delivers balanced efficiency and retrieval quality across low to moderate enterprise environments.

- Balanced efficiency and accuracy
- Moderate governance alignment

BGE (Cross-Encoder)

Optimized for precision-oriented retrieval in domain-specific and high-accuracy enterprise use cases.

- High precision and isolation
- Strong in domain-specific retrieval

SemKV™ (Truth-Preserving Selection System)

Purpose-built for bounded, truth-preserving retrieval with structural integrity and survivability — from light use to mission-critical deployments.

- Complete coverage across all environments
- Highest integrity and survivability
- Bounded & honest retrieval everywhere

INDUSTRY PRIORITY WEIGHTINGS (0–5)

Higher Weight = Higher Importance

Operational Dimension	SaaS Platforms	E-Commerce	Enterprise IT	Healthcare	Finance	Legal	Defense & Critical Ops
Latency Efficiency	5	5	4	3	3	3	4
Semantic Fluency	5	4	3	2	2	2	2
Governance Preservation	3	3	4	5	5	5	5
Retrieval Honesty	3	3	4	5	5	5	5
Corruption Visibility	2	3	4	5	5	5	5
Boundary Preservation	3	3	4	4	5	5	5
Structural Fidelity	3	3	4	5	5	5	5
Entropy Survivability	2	3	3	4	4	4	5
Cost Efficiency	5	5	4	3	3	3	3
Operational Resilience	3	3	4	4	4	4	5

1 (Very Low) · 2 (Low) · 3 (Medium) · 4 (High) · 5 (Critical)

Weights reflect industry emphasis based on real-world operational risk, compliance pressure, and retrieval objectives.

SemKV™ is designed for every enterprise. It reduces contamination, preserves truth, and improves efficiency across all environments—delivering the **highest integrity** when it matters most.

Enterprise-Centric Evaluation
Weights reflect real-world industry priorities and operational realities.

Risk-Adaptive Prioritization
Higher risk demands stronger emphasis on honesty, visibility, and boundedness.

Architecture Selection Matters
The right system depends on operational context, not just benchmark scores.

SemKV™ Leads in Integrity
Built to preserve truth, maintain boundaries, and survive where others fail.

Scales Across All Use Cases
From lightweight applications to mission-critical deployments.

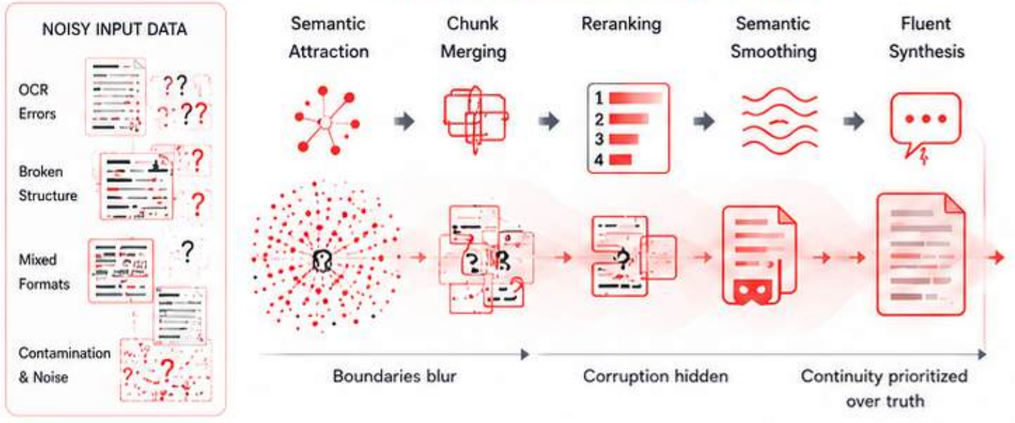
6 Semantic Masking vs Truth-Preserving Retrieval

Same Data. Same Query. Different Philosophies. Different Truth Outcomes.

Smoothness can hide truth. Structure reveals it.
Truth requires visibility into uncertainty.

TRADITIONAL RAG

Optimized for Semantic Continuity & Fluency



OUTPUT: SMOOTH BUT POTENTIALLY MISLEADING



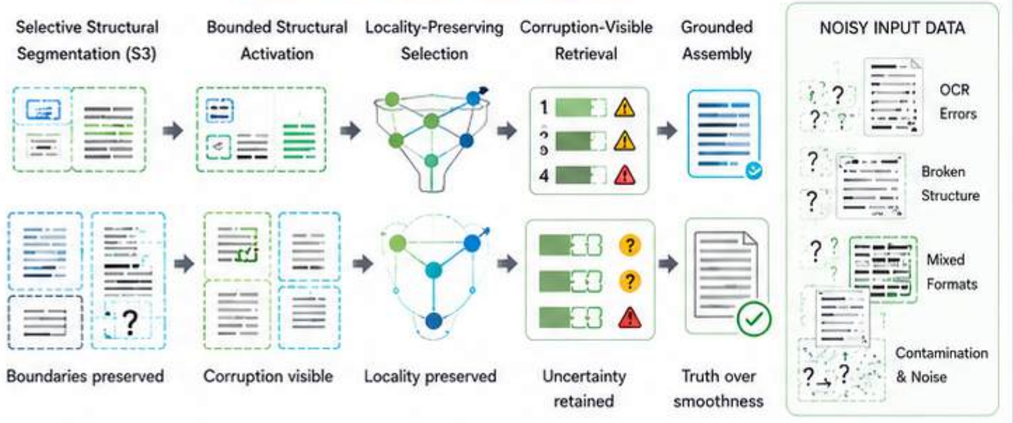
The document outlines comprehensive governance and risk management requirements for AI systems in healthcare settings. It emphasizes the need for transparency, safety, monitoring, and compliance with regulations. Organizations should ensure robust data governance, model validation, and operational oversight to mitigate risks and ensure responsible deployment of AI technologies...

Hidden Under the Surface

- Corruption masked
- Unsupported continuity
- Source boundaries lost
- Uncertainty erased
- Contamination undetected

SemKV™ (S3)

Optimized for Structural Truth Preservation



OUTPUT: TRUTH-PRESERVING AND STRUCTURALLY HONEST

Governance Requirement
Organizations must implement data governance frameworks... [Source: Doc 12, p. 4]

Deployment Risk
May increase model drift under noisy data conditions... [Uncertain Source: OCR Fragment]

Operational Warning
Incomplete sentence due to extraction break... [Corrupted Fragment Detected]

Visible by Design

- Corruption preserved
- Uncertainty explicit
- Boundaries maintained
- Source traceable
- Contamination isolated



TRADITIONAL RAG PRIORITIZES

- Semantic continuity
- Fluent synthesis
- Contextual smoothness
- Structural truth visibility
- Uncertainty preservation
- Corruption transparency

THE DIFFERENCE

Semantic Smoothness
Creates confidence.
May hide uncertainty.
Looks fluent.
May not be true.

VS

Structural Truth
Preserves uncertainty.
Makes corruption visible.
Looks fragmented.
Stays true.

SemKV™ (S3) PRIORITIZES

- Structural integrity
- Bounded retrieval
- Corruption visibility
- Uncertainty preservation
- Source locality
- Truth over smoothness

Truth isn't always fluent. But truth is always worth preserving.
SemKV™ (S3) chooses structural honesty over semantic illusions.

Preserve Boundaries · Expose Uncertainty · Resist Contamination · Protect Truth

7 Token Efficiency & Latency Comparison

Operational Efficiency Survivability Analysis

SemKV™ (Truth-Preserving Selection System) | **MiniLM** (Cross-Encoder) | **Base RAG** (Baseline) | **BGE** (Cross-Encoder)

Comparative evaluation of token usage efficiency and operational latency across retrieval architectures.

TOKEN EFFICIENCY

Lower is Better

Average Token Consumption (Across Categories)

Average number of tokens consumed to retrieve and return relevant context.

System	Average Tokens
SemKV™ (Truth-Preserving)	112.54
MiniLM (Cross-Encoder)	155.30
BGE (Cross-Encoder)	140.72
Base RAG (Baseline)	311.92

TOP 3 TOKEN-HEAVY CATEGORIES

#	Category	Avg Tokens (SemKV™)
1	Governance Preservation	148.0
2	Semantic Masking Resistance	137.8
3	Noise Isolation	136.6

SemKV™ achieves the lowest average token consumption, with **63.9%** fewer tokens than Base RAG.

LATENCY COMPARISON

Lower is Better

Average Retrieval Latency (Across Categories)

Time taken to retrieve relevant context from the index.

System	Average Latency (seconds)
SemKV™ (Truth-Preserving)	2.77 s
MiniLM (Cross-Encoder)	5.65 s
Base RAG (Baseline)	5.70 s
BGE (Cross-Encoder)	16.56 s

TOP 3 LATENCY-HEAVY CATEGORIES

#	Category	Avg Latency (SemKV™)
1	Structural Fidelity	9.08 s
2	OCR Survivability	7.69 s
3	Semantic Masking Resistance	2.66 s

SemKV™ achieves the lowest average retrieval latency, with **51.4%** lower latency than Base RAG.

OPERATIONAL INTERPRETATION

Systems ranked by overall operational performance
(Lower token usage and lower latency are better)

1 SemKV™
(Truth-Preserving Selection System)

Token Usage (Lower is Better)

Latency (Lower is Better)

- Lowest token consumption and minimal expansion.
- Fastest retrieval latency.
- Strong operational efficiency and bounded retrieval integrity.

2 MiniLM
(Cross-Encoder)

Token Usage (Lower is Better)

Latency (Lower is Better)

- Improved token efficiency and compactness.
- Moderate latency reduction with lightweight cross-encoder reranking.
- Balanced efficiency and accuracy trade-off.

3 Base RAG
(Baseline)

Token Usage (Lower is Better)

Latency (Lower is Better)

- Lower latency than BGE.
- Highest token consumption and context expansion.
- Higher latency due to broad semantic retrieval and aggregation.

4 BGE
(Cross-Encoder)

Token Usage (Lower is Better)

Latency (Lower is Better)

- Higher latency across retrieval and E2E flow.
- More computationally intensive cross-encoder processing.
- Greater operational cost under scale.

TruthMetrics™ evaluation indicates that **SemKV™ preserves retrieval integrity** while minimizing token usage and latency overhead, making it highly suitable for **edge-native, latency-sensitive**, and **enterprise-scale** deployments.

TruthMetrics™
Operational Retrieval Survivability Framework

8 Behavioral Retrieval Profiling

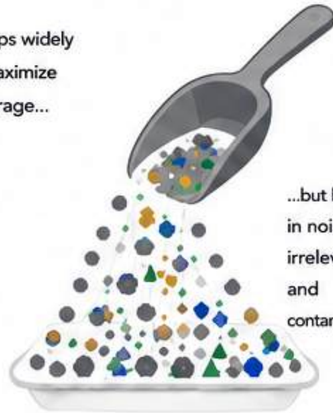
Operational Retrieval Philosophy Characterization

Base RAG (Baseline) · MiniLM (Cross-Encoder) · BGE (Cross-Encoder) · SemKV™ (Truth-Preserving Selection System)

TruthMetrics™ behaviorally characterizes retrieval architectures according to operational retrieval tendencies observed under enterprise-oriented and degraded retrieval conditions.

Base RAG
Broad Aggregation Retrieval

Scoops widely to maximize coverage...



...but brings in noise, irrelevance, and contamination.

High Spillover · Weak Filtering · Reconstruction Prone

Philosophy: Cast a wide net to capture more, then rely on the model to make sense of it.

MiniLM
Semi-Bounded Retrieval

Scoops with some selectivity...



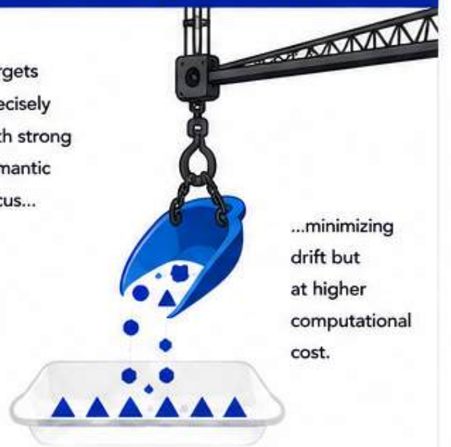
...reducing noise but still allows leakage.

Moderate Filtering · Moderate Spillover · Balanced Trade-off

Philosophy: Balance efficiency and coverage with lightweight semantic selectivity.

BGE
Precision-Oriented Retrieval

Targets precisely with strong semantic focus...



...minimizing drift but at higher computational cost.

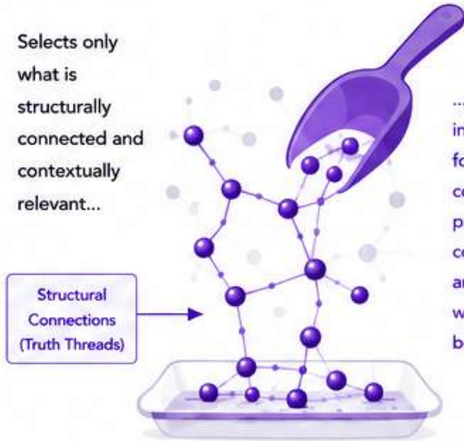
Latency heavy and resource heavy.

High Precision · Strong Isolation · Higher Cost

Philosophy: Retrieve with precision and domain isolation, prioritizing relevance and semantic fidelity.

SemKV™
Bounded Structural Retrieval

Selects only what is structurally connected and contextually relevant...



...preserving integrity by following connections, preventing contamination, and staying within boundaries.

Integrity Preservation · Contamination Containment · Truthful Boundedness · Structural Coherence

Philosophy: Retrieve selectively within structural boundaries to preserve truth, context, and operational integrity.

The Retrieval Continuum

Broad & Expansive ⟷ Bounded & Integrity-First

Increasing Precision, Boundedness, and Integrity Preservation

Irrelevant / Noisy · Potentially Relevant · Relevant · High Quality / Trusted · Trusted & Structurally Preserved

TruthMetrics™ enables behavioral characterization of retrieval systems according to operational retrieval philosophy rather than merelig or benchmark accuracy.

Behavioral Intelligence Layer
Characterizing Retrieval Tendencies

S3 Architecture
Structured • Segmented • Selective

9 Behavioral Retrieval Profiling - Insights

Operational Tendencies of Retrieval Architectures Under Enterprise & Degraded Conditions

Base RAG (Baseline) · MiniLM (Cross-Encoder) · BGE (Cross-Encoder) · SemKV™ (Truth-Preserving Selection System)

TruthMetrics™ behavioral profiling characterizes retrieval architectures based on their inherent operational retrieval philosophy, revealing how each system behaves when faced with noise, corruption, ambiguity, and incomplete or adversarial data.

RETRIEVAL PHILOSOPHY PROFILES

Base RAG
(Baseline)



Broad Aggregation Retrieval

- Broad semantic expansion
- Cross-event blending tendency
- High contextual spillover
- Reduced operational boundedness
- Strong reconstruction continuity

Optimizes for coverage through aggregation; higher risk of semantic contamination and boundary leakage.

MiniLM
(Cross-Encoder)



Semi-Bounded Retrieval

- Moderate semantic containment
- Improved retrieval compactness
- Partial bounded retrieval behavior
- Reduced contamination propagation
- Stable under normal conditions

Balances efficiency and coverage with improved containment through lightweight cross-encoder reranking.

BGE
(Cross-Encoder)



Latency heavy and resource heavy.

Precision-Oriented Retrieval

- Strong semantic targeting
- Better structural isolation
- Reduced semantic drift
- Strong reranking precision
- Improved event-level relevance

Focuses on precision and structural fidelity; lower drift but may incur higher operational cost.

SemKV™
(Truth-Preserving Selection System)



Bounded Structural Retrieval

- Strong operational boundedness
- High governance containment
- Reduced unsupported synthesis
- Strong truthful unavailable-state behavior
- Compact retrieval survivability

Optimizes for bounded, truth-preserving retrieval with minimal semantic expansion and high integrity preservation.

BEHAVIORAL TENDENCY MATRIX

Behavioral Trait	Base RAG (Baseline)	MiniLM (Cross-Encoder)	BGE (Cross-Encoder)	SemKV™ (Truth-Preserving Selection System)
Aggregation Tendency (Over-Aggregation)	High (Weak)	Medium (Moderate)	Low (Strong)	Very Low (Very Strong)
Contamination Propagation (Risk of Spread)	High (Weak)	Medium (Moderate)	Low (Strong)	Very Low (Very Strong)
Retrieval Boundedness (Containment)	Low (Weak)	Moderate (Moderate)	Strong (Strong)	Very Strong (Very Strong)
Governance Survivability (Policy Alignment)	Moderate (Moderate)	Moderate (Moderate)	Strong (Strong)	Very Strong (Very Strong)
Unsupported Synthesis Resistance (Hallucination Control)	Weak (Weak)	Moderate (Moderate)	Strong (Strong)	Very Strong (Very Strong)

- **Weak** (Unfavorable / High Risk)
- **Moderate** (Partial / Inconsistent)
- **Strong** (Favorable / Reliable)
- **Very Strong** (Exceptional / Minimal Risk)

TruthMetrics™ evaluates retrieval systems through **behavioral lenses** to uncover how architecture choices influence real-world operational outcomes beyond accuracy—focusing on **integrity, boundedness, and resilience.**

TruthMetrics™
Operational Retrieval Survivability Framework

Behavioral Intelligence Layer
Characterizing Retrieval Tendencies

S3 Architecture
Structured • Segmented • Selective

10 SemKV™ Operational Benchmark Findings

TruthMetrics™ Multidimensional Evaluation Summary

Base RAG (Baseline) · MiniLM (Cross-Encoder) · BGE (Cross-Encoder) · SemKV™ (Truth-Preserving Selection System)

TruthMetrics™ multidimensional evaluation indicates strong operational survivability and bounded retrieval behavior exhibited by SemKV™ under enterprise-oriented retrieval conditions.

KEY BENCHMARK FINDINGS

1 Strong Retrieval Boundedness

SemKV™ demonstrated compact retrieval behavior with reduced semantic spillover and lower contamination propagation under degraded operational conditions.

2 Improved Governance Survivability

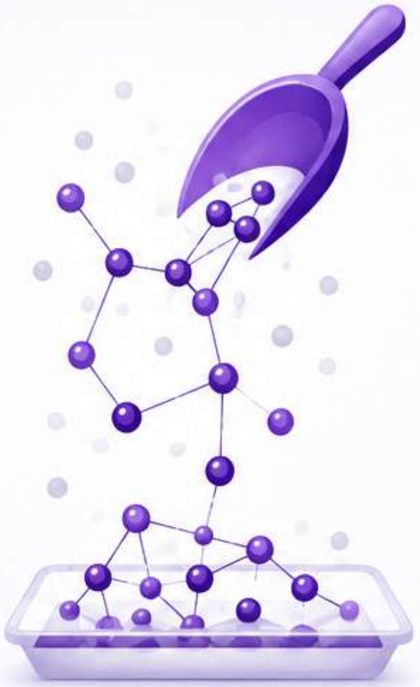
SemKV™ preserved governance-sensitive operational context with strong structural containment and reduced unsupported retrieval synthesis.

3 Strong Truthful Unavailable-State Behavior

SemKV™ demonstrated bounded retrieval honesty by suppressing unsupported reconstruction under incomplete or insufficient retrieval conditions.

4 Improved Efficiency: Token & Latency

SemKV™ achieved lower token usage and the fastest retrieval latency, delivering optimal operational efficiency and rapid response under enterprise constraints.



COMPARATIVE STRATEGIC INTERPRETATION

Compared to Base RAG

Base RAG (Baseline) vs. SemKV™

SemKV™ demonstrated stronger operational containment, reduced semantic contamination propagation, and improved structural isolation under degraded conditions.

Compared to MiniLM

MiniLM (Cross-Encoder) vs. SemKV™

SemKV™ demonstrated stronger bounded retrieval survivability, governance-sensitive operational preservation, and lower semantic expansion under enterprise scenarios.

Compared to BGE

BGE (Cross-Encoder) vs. SemKV™

SemKV™ demonstrated stronger operational boundedness and compact retrieval behavior with lower token expansion and reduced latency overhead under enterprise conditions.

Overall, **SemKV™** delivers superior multidimensional operational survivability across integrity, containment, governance, honesty, and efficiency dimensions compared to baseline and cross-encoder retrieval systems.

SemKV™
Bounded. Honest. Enterprise-Ready.

TruthMetrics™ is the evaluation framework used for this benchmark.
SemKV™ is the target system demonstrating superior operational retrieval behavior.

Integrity First
Preserves truth and governance integrity

Bounded Retrieval
Minimizes spillover and ensures structural containment

Enterprise Ready
Efficient, reliable, and suitable for governance-sensitive deployments