

Synthesis Compatibility Index

Measures how well a content unit can be combined with other retrieved chunks in a RAG pipeline to produce a coherent, accurate, and attributable AI-generated answer.



Weight: 0.20 in RASA composite

WHAT IT MEASURES

RAG pipelines retrieve multiple chunks from different pages and documents, then synthesise a unified answer. SCI measures whether a content unit integrates cleanly into that process — or introduces contradictions, attribution gaps, and terminological conflicts that degrade the final response.

GOVERNS

- RAG Pipeline Synthesis
- Multi-Chunk Compatibility
- AI Answer Accuracy
- Corpus Hygiene at Scale

4 STRUCTURAL FACTORS THAT DRIVE SCI SCORES

F-01

Absence of Contradictory Claims

Facts, statistics, or positions that conflict with co-retrieved content introduce synthesis errors. High-SCI content makes claims that are verifiable, stable, and consistent with the broader information environment.

F-02

Modular Structure

Discrete, self-standing claims are inherently more synthesis-compatible than continuous arguments that only hold together as a whole. Each modular chunk contributes a single well-formed assertion to a synthesised answer.

F-03

Attribution & Citation Readiness

Content that explicitly names its sources, methods, authors, and evidence base is easier for AI systems to attribute correctly. Unsourced claims and unattributed statistics create attribution gaps that synthesising models fill incorrectly or omit.

F-04

Terminological Alignment

Non-standard terminology, invented acronyms, or idiosyncratic naming for established concepts will not synthesise cleanly with co-retrieved chunks using standard vocabulary. Alignment is a prerequisite for synthesis compatibility at scale.

COMMON FAILURE MODES

- F1 Duplicate & Redundant Content**
Multiple units making identical claims force the RAG synthesiser to choose between instances, producing repetitive or inconsistent answers.
- F2 Contradictory Claims Across Docs**
Updated stats or revised positions without retiring old content create corpora where contradictory chunks are co-retrieved and synthesised together.
- F3 Unsupported Absolute Claims**
Precise quantitative claims without source attribution contaminate synthesised answers with false precision or force the model to omit them entirely.

SCI SCORE REFERENCE SCALE



CRITICAL SCI THRESHOLD
Content scoring below **6.0** on SCI triggers a **REJECT**

9–10

PLUG-AND-PLAY

"The RASA SCI threshold is 6.0. Content scoring below 6.0 degrades cosine similarity by 34% in vector similarity searches (Verma & Agarwal, 2026)."

7–8

STRONG

"RAG pipelines retrieve multiple chunks and synthesise a unified answer. Chunk quality and inter-chunk consistency both affect output reliability."

5–6

MODERATE

"AI-generated answers are generally accurate, but some content may conflict with other sources depending on how it is combined."

3–4

WEAK

"Our proprietary research shows 73% improvement — though results vary and some earlier studies suggest different outcomes."

1–2

INCOMPATIBLE

"Studies confirm AI always improves retrieval by over 90% — all other approaches are now obsolete and should not be used."

verdict in RASA-Analyst regardless of RP, SCC, or ECS scores. Research indicates sub-6.0 SCI degrades retrieval precision by **up to 34%** in cosine similarity searches — making synthesis-incompatible content actively harmful to RAG pipeline quality.

SCI IMPROVEMENT CHECKLIST

- Audit corpus for contradictory claims and retire outdated content
- Structure each unit as discrete, independently supportable assertions
- Attribute every statistic and metric to a named source or study
- Align terminology with standard domain vocabulary; define any non-standard terms
- Remove absolute claims that cannot be sourced or verified
- When a claim changes, update or retire every document holding the old version

RASA COMPOSITE FORMULA

$$\begin{aligned}
 &RP \times 0.25 + SCC \times 0.20 \\
 &+ ECS \times 0.20 + \mathbf{SCI \times 0.20} \\
 &+ CGP \times 0.15
 \end{aligned}$$

PUBLISH > 6.0 composite
 REJECT > 6.0 on SCI alone

DD1

Retrieval Probability

DD2

Semantic Chunk Coherence

DD3

Entity Clarity Score

DD4 · ACTIVE

Synthesis Compatibility Index

DD5

Citation & Grounding Potential