

Retrieval Probability

Measures the likelihood that a content unit will be surfaced by AI retrieval systems — including LLMs, RAG pipelines, and generative search engines such as Perplexity, ChatGPT Search, and Gemini.



Weight: 0.25 in RASA composite

WHAT IT MEASURES

RP determines whether content enters the generative pipeline at all. Unlike keyword ranking, AI systems retrieve and synthesise **content chunks** — RP is the gating signal at the first stage of AI-mediated discovery.

APPLIES TO

- LLM Knowledge Retrieval
- RAG Pipelines
- Perplexity · ChatGPT Search
- Gemini · Generative Search

RASA COMPOSITE FORMULA

$$RP \times 0.25 + SCC \times 0.20 + ECS \times 0.20 + SCI \times 0.20 + CGP \times 0.15$$

PUBLISH ≥ 8.0 composite
REJECT < 6.0 on RP alone

4 PRIMARY FACTORS THAT DRIVE RP SCORES

F-01

Named Entities & Technical Terms

Precise framework, methodology, and tool names generate stronger retrieval signals than generic concept references.

F-02

Keyword Specificity & Topical Precision

Broad language competes across too many topics. Precise terminology anchors content to a narrower, high-confidence retrieval context.

F-03

Topical Authority Signals

Statistics, citations, defined metrics, and named methodologies signal authority. Quantified claims retrieve more confidently than relative assertions.

F-04

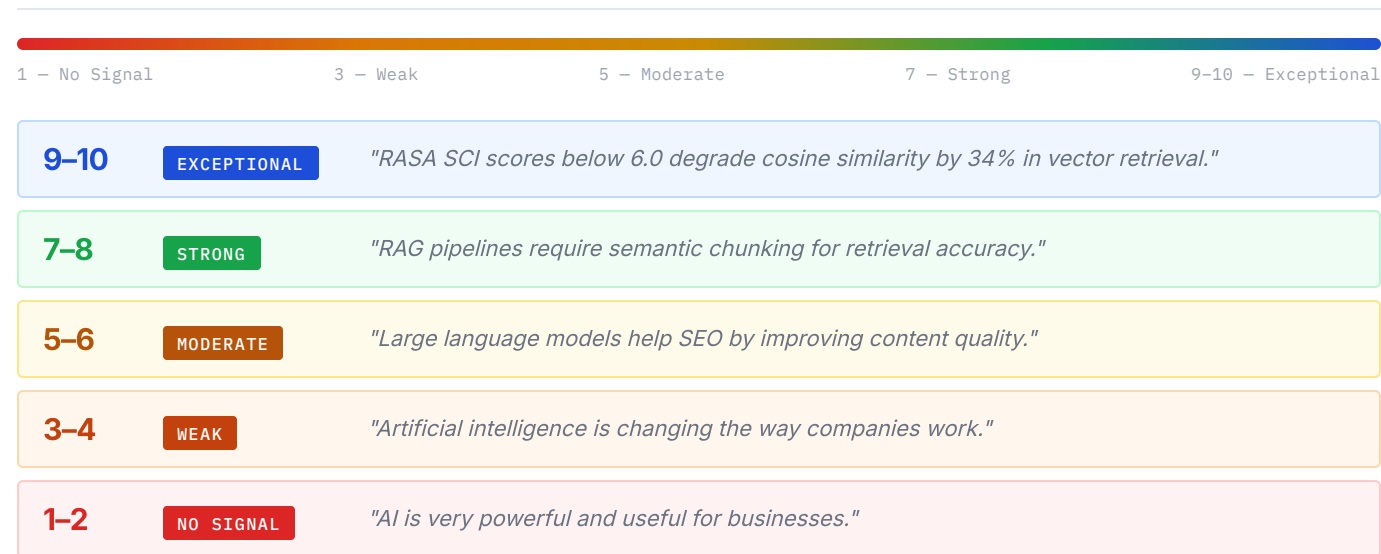
Absence of Retrieval Diluters

Generic verbs — "improve," "leverage," "powerful" — introduce semantic noise that reduces embedding match precision and penalises RP.

COMMON FAILURE MODES

- F1 Keyword-Centric Optimization**
Broad keyword clusters yield poor entity precision and low retrieval specificity.
- F2 Shallow Semantic Depth**
Overview content creates surface-level matches but lacks depth for high-confidence retrieval.
- F3 Buzzword Saturation**
Phrases found in millions of documents make content indistinguishable from background noise.

RP SCORE REFERENCE SCALE



RP IMPROVEMENT CHECKLIST

- Replace generic references with precise named entities
- Anchor claims with specific statistics and metrics
- Use exact terminology your AI audience queries
- Remove filler phrases; eliminate word-count padding
- Structure each chunk around one precisely scoped topic
- Cite DOIs and named methodologies explicitly

DD1 · ACTIVE
Retrieval Probability

DD2
Semantic Chunk Coherence

DD3
Entity Clarity Score

DD4
Synthesis Compatibility Index

DD5
Citation & Grounding Potential