RASA FRAMEWORK · DISCOVERABILITY DYNAMIC 1 OF 4

# 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.

**RP**
1 – 10

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 × 0.25** + SCC × 0.20
+ ECS × 0.20 + SCI × 0.20
+ CGP × 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.

## RP SCORE REFERENCE SCALE

1 — No Signal · 3 — Weak · 5 — Moderate · 7 — Strong · 9–10 — Exceptional

| | | |
|---|---|---|
| 9–10 | EXCEPTIONAL | *"RASA SCI scores below 6.0 degrade cosine similarity by 34% in vector retrieval."* |
| 7–8 | STRONG | *"RAG pipelines require semantic chunking for retrieval accuracy."* |
| 5–6 | MODERATE | *"Large language models help SEO by improving content quality."* |
| 3–4 | WEAK | *"Artificial intelligence is changing the way companies work."* |
| 1–2 | NO SIGNAL | *"AI is very powerful and useful for businesses."* |

## 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 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




Verma, A. & Agarwal, S. (2026). Retrieval-Aware Semantic Architectures (RASA). Nebula Personalization Tech Solutions Pvt. Ltd. · DOI: 10.5281/zenodo.20325460