

EcoSearch

Evidence-first Retrieval for Transparent Question Answering

EcoSearch is a retrieval-first system for transparent evidence access over prepared corpora. Instead of generating answers first and looking for evidence later, EcoSearch retrieves supporting passages first, keeps the evidence visible, and helps users inspect whether the retrieved text is strong enough to answer.

The goal is simple: make document-based question answering more inspectable, more grounded, and easier to evaluate before any external generation layer is involved.

The problem

Many chatbot-first QA systems feel fluent even when the evidence is weak, missing, or hard to inspect. In document workflows, this creates a practical problem: users need to know not only what an answer might be, but *where the answer comes from*.

For retrieval-augmented workflows, the key risk is not only hallucination. A more basic failure often happens earlier: the system retrieves a passage that sounds related but does not actually contain the evidence needed to answer the user's question.

EcoSearch focuses on that layer: retrieval quality, evidence visibility, and answer containment.

The EcoSearch approach

EcoSearch is built around an evidence-first workflow:

1. Prepare a corpus from source documents.
2. Split it into overlapping evidence chunks.
3. Retrieve candidate passages using semantic and lexical signals.
4. Rerank and score the retrieved evidence.
5. Show the supporting passage, sentence references, and retrieval metadata.
6. Let the user inspect whether the retrieved evidence is sufficient.

This makes the system useful as an evidence access tool, but also as a way to inspect retrieval behaviour, compare scoring strategies, and evaluate whether retrieved passages actually contain the evidence needed by the query.

What EcoSearch does

EcoSearch combines several retrieval and evaluation components:

Semantic retrieval

Dense embeddings are used to identify passages that are semantically related to the user's question.

Lexical and anchor evidence

Rare or exact textual anchors help preserve important details that pure semantic similarity may blur.

Reranking

Candidate passages can be reranked to improve the ordering of retrieved evidence.

Confidence routing

The system can distinguish stronger matches from weaker or ambiguous ones, supporting actions such as access the top passage, compare alternatives, or clarify the query.

Evidence visibility

Retrieved chunks, sentence indices, summaries, and scores remain visible, so users can inspect the basis for the result.

Mobile preview

EcoSearch has been extended into a standalone mobile preview for iOS and Android.

The mobile app supports local-first retrieval on prepared corpora. Corpus preparation can still use an OpenAI key, but once a corpus is prepared, the core retrieval workflow can run locally inside the app.

Maurizio Scibilia

AI Engineer | Retrieval, RAG Evaluation & Agentic AI Review

Italy (Remote) | [mail](#) | [LinkedIn](#) | [website](#) | +39 392 182 6406

This makes EcoSearch useful as a lightweight evidence-access tool for document collections where users want retrieval, inspection, and source-grounded navigation without relying on a chatbot-first interface.

Evaluation

EcoSearch v1.1 focuses on retrieval quality across multiple corpora.

The evaluation work does not rely only on whether a retrieved passage is semantically similar to the query. Instead, it asks a stricter question:

Does the retrieved passage actually contain the evidence needed to answer the question?

Current evaluation work includes:

Oracle-overlap analysis

An internal evaluation harness compares retrieved passages against oracle evidence spans generated for the experiment. This is not a standard user-facing feature of the app; it is a research and validation method used to measure whether EcoSearch retrieves answer-containing text.

Focus-hit diagnostics

A lighter diagnostic checks whether retrieval reaches the focused evidence needed for the query.

Semantic/lexical balance experiments

EcoSearch tests different semantic and lexical weighting ratios to understand when dense semantic retrieval needs support from rare or exact-anchor evidence.

These experiments are being documented in the EcoSearch v1.1 white paper.

Why it matters

EcoSearch is not designed to compete with general-purpose chatbots. It focuses on a narrower and more testable layer: retrieval quality, source visibility, answer containment, and grounded evidence access. This matters for workflows where the user needs to inspect supporting text before relying on an answer or passing evidence to an external generation layer, especially in research, documentation, compliance, education, internal knowledge bases, and domain-specific QA. The first question is always:

What evidence did the system retrieve, and is it good enough?

Current status

EcoSearch currently includes:

- Retrieval-first web prototype
- Standalone mobile preview for iOS and Android
- Local-first retrieval workflow on prepared corpora
- Semantic/lexical balance experiments
- Internal oracle-overlap and focus-hit evaluation
- EcoSearch v1.1 white paper in progress

A full review package is available on request.

Review package

The EcoSearch review package can include:

- Mobile preview access
- EcoSearch v1.1 white paper draft
- Short instruction booklet
- Demo script and suggested test queries
- Project profile and technical notes

Access is shared selectively to avoid unintended LLM costs, dataset leakage, and uncontrolled demo usage.

Maurizio Scibilia

AI Engineer | Retrieval, RAG Evaluation & Agentic AI Review

Italy (Remote) | [mail](#) | [LinkedIn](#) | [website](#) | +39 392 182 6406