

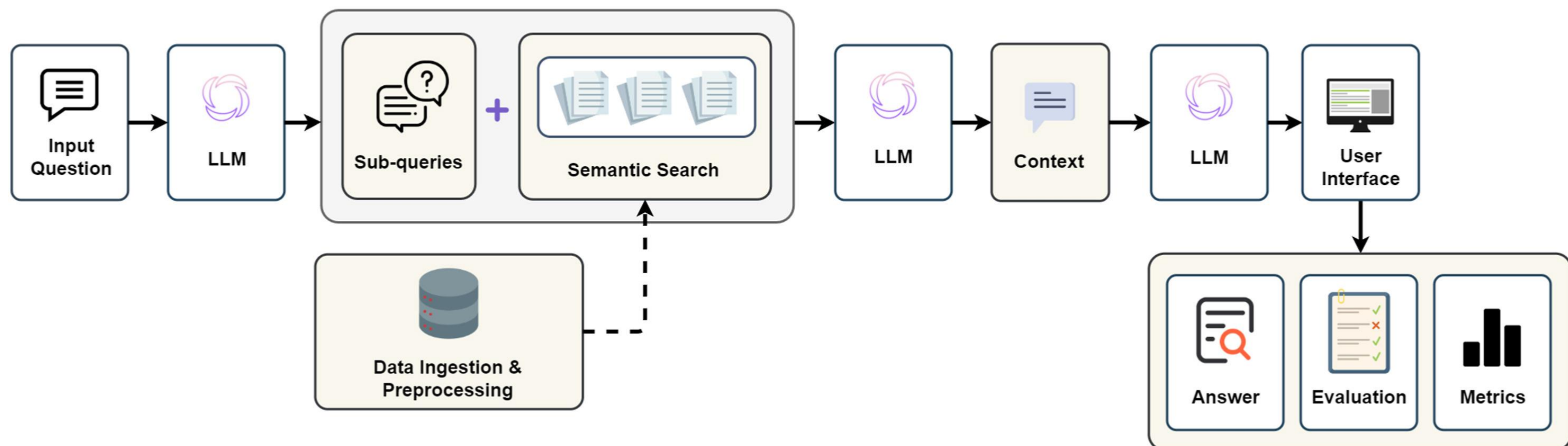
Background

- Craniofacial surgery spans syndromic craniosynostosis to complex adult reconstruction, demanding integration of three-dimensional anatomy, growth-dependent timing, and multidisciplinary care across plastic surgery, neurosurgery, otolaryngology, pediatrics, and genetics.
- General-purpose large language models (LLMs) offer rapid reasoning but rely on static training data and can generate non-verifiable ("hallucinated") content, unacceptable risks in surgical decision-making.
- Retrieval-augmented generation (RAG)** grounds model outputs in curated knowledge and enables **direct citation of supporting evidence**, a prerequisite for defensible clinical recommendations.
- No domain-specific, citation-enabled AI system has been developed for craniofacial surgery.

Objectives

- Develop **CASPER (Craniofacial AI Surgical Planning Evidence Retriever)** — the first specialty-specific, citation-enabled, multi-modal RAG system for craniofacial surgery.
- Construct a curated craniofacial literature knowledge base using **RAPTOR hierarchical architecture**.
- Evaluate CASPER across 25 clinical questions spanning craniosynostosis, cleft care, trauma, reconstruction, pediatric airway, aesthetic surgery, oncology, and emerging technologies.
- Demonstrate the feasibility of **transparent, evidence-grounded AI decision support** for complex surgical planning and education.

Question Answering: Detailed View



Methods

Knowledge Base

- 8,561 open-access, peer-reviewed craniofacial articles
- PubMed Central, 2000–2025
- Terms span cleft lip/palate, craniosynostosis, orthognathic, orbital reconstruction, Le Fort osteotomy, facial feminization/masculinization, and maxillofacial/skull reconstruction.
- Cleaned text segmented into 1,000-token chunks with 200-token overlap.

Architecture

- RAPTOR (Recursive Abstractive Processing for Tree-Organized Retrieval).
- 1,536-dim embeddings (text-embedding-ada-002) clustered via UMAP + Gaussian Mixture Models; stored in Chroma vector database.
- Multi-modal: text and image ingestion.

Query Processing & Evaluation

- Each query decomposed into up to 5 sub-queries; top-3 segments retrieved per sub-query; synthesized by GPT-5 with explicit citations.
- Performance assessed by semantic similarity (SEM-eval), manual content-coverage review, and manual citation accuracy verification across 25 clinical questions.

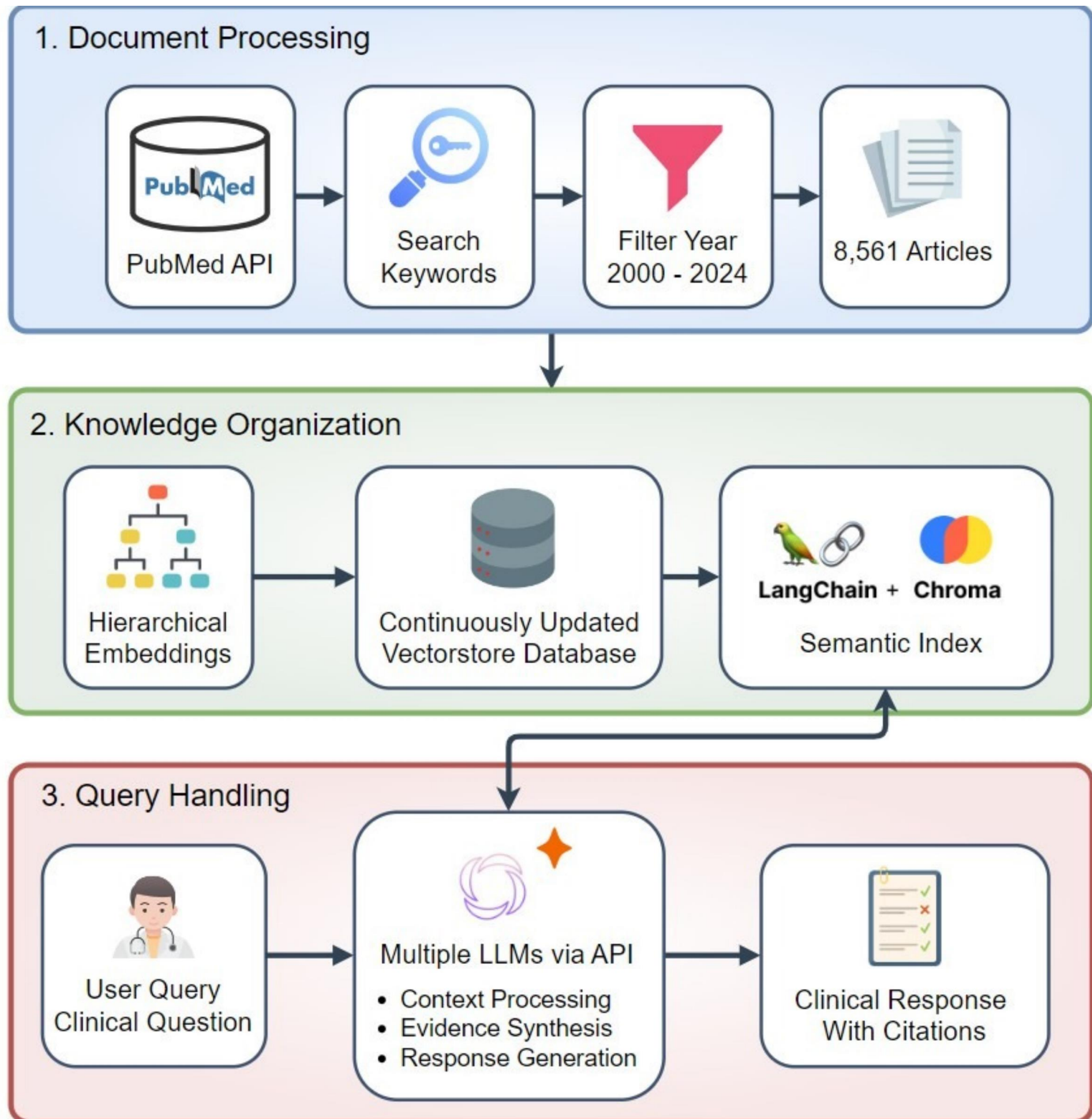


Figure 1. CASPER three-module architecture: document processing, knowledge organization, and query handling.

Outcomes

0.89

mean SEM-eval
(range 0.81–0.96)

8,561

peer-reviewed articles
in knowledge base

25

clinical questions
across subspecialties

7.8

sources synthesized
per query (avg)

Top-performing clinical domains

- Oncologic surgery 0.95 | Pediatric airway 0.93 ± 0.04 | Facial trauma 0.93 ± 0.03 .
- Top questions: **Pierre Robin MDO (0.96)**, **orbital floor fracture (0.95)**, **craniofacial resection (0.95)**.

Evidence integration scaled with complexity

- Focused scenarios (3–6 docs): 0.92 ± 0.04 . Complex multi-domain questions (10–11 docs): 0.87 ± 0.03 .

Qualitative review confirmed transparency

- High content coverage** across nearly all domains; referenced sources consistently **relevant and directly supportive** of system claims.

Lower scores in emerging / nuanced areas

- Le Fort III contraindications (0.81), facial feminization (0.81), midface-distraction relapse (0.82), plagiocephaly vs lambdoid synostosis (0.83) - reflecting evolving evidence bases and multi-step reasoning demands.

Multi-modal prototype validated

- CASPER integrated craniofacial image analysis with literature retrieval to generate a prioritized differential diagnosis (e.g., Crouzon syndrome as leading diagnosis from a pediatric craniofacial photograph).

Conclusion

- CASPER is the **first retrieval-augmented, citation-enabled AI clinical decision support system** tailored to craniofacial surgery.
- Transparent sourcing and citation traceability directly address LLM hallucination risk and enhance clinical trust.
- Dual **clinical + educational utility**: supports real-time surgical planning and structured, evidence-based learning for trainees.
- Multi-modal architecture positions CASPER for workflows where visual interpretation is central, cephalometric analysis, cranial deformity assessment, and osteotomy planning.