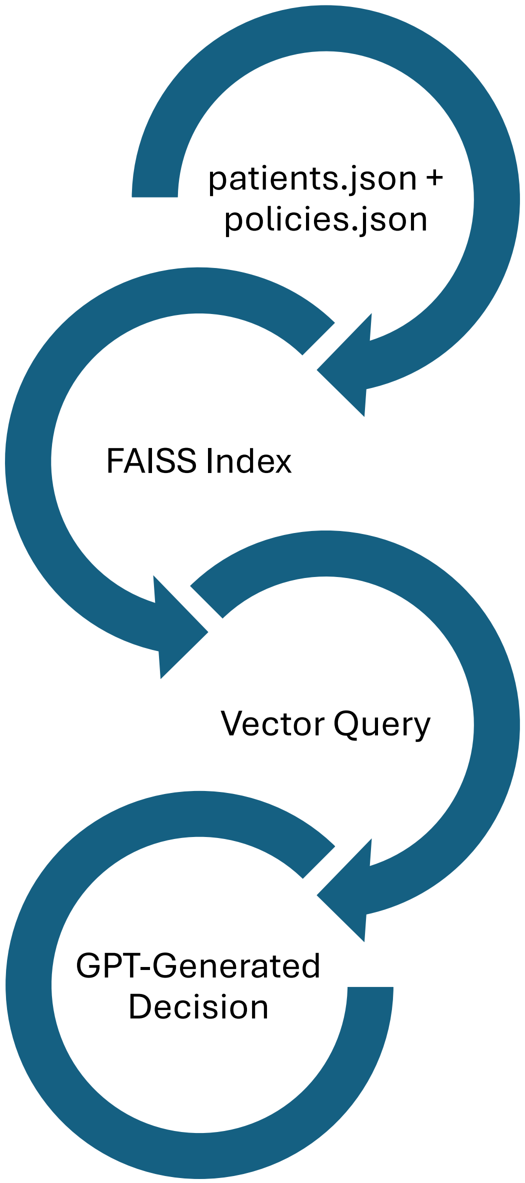
AI-Powered Prior Authorization with RAG

Rayan Elias

AI-Powered Prior Authorization with RAG

- Core Process:
    1. Embed structured data (patients, policies) into vector representations
    2. Store embeddings in FAISS for fast similarity search
    3. Retrieve relevant policy based on patient data
    4. Generate an AI-powered authorization decision
- Outcome: Faster and more accurate prior authorization decisions, reducing manual workload



System Architecture Overview

Data Processing

- Flattening JSON: Converted patients.json and policies.json into structured text for embeddings.
- Handled Variable Data Structures: Supported metadata-heavy and simple policy formats.

Embedding & Storage

- Generated OpenAI Embeddings: Converted patient & policy text into vector representations.
- FAISS Indexing: Stored embeddings for fast, scalable similarity searches.

Retrieval & Response Generation

- Efficient Querying: Compared patient data against policy database to find best-matching rules.
- AI-Assisted Decision Support: Used GPT to generate explanatory responses for approvals/rejections.
- Ensured Accuracy: AI responses are grounded in actual policy data to prevent hallucinations.

Example: Captain Hook

Query Request: To determine if Captain Hook meets the criteria for approval of an epidural steroid injection (ESI) under the provided policy, we need to consider several factors outlined in the InterQual® 2022 policy and the criteria for coverage determinations

Example Vector Comparison: The policy limits the frequency and duration of ESIs. Without knowing how many injections, if any, Captain Hook has had, it's challenging to approve another ESI per the limitations (Note 29)

- Criteria 29: “ESIs are limited to a maximum of four (4) sessions per spinal region in a rolling twelve (12) month period (1)”

Example Vector Comparison: Considering his medication list, potential issues like uncontrolled diabetes mellitus could be a concern due to insulin-potentiation with steroids (Note 24). Without specific laboratory values or a comprehensive health profile, this remains speculative

- Criteria 24: “...Uncontrolled diabetes mellitus- Insulin-dependent diabetics are at risk of elevated blood sugars after steroid injections...”

Based on the available information, Captain Hook does not currently meet the criteria for approval of an ESI according to the listed policy. Key information missing as demonstrated previously. Contradictions in his patient profile with the policy reject this case.

Why Use Embeddings & Vectors?

Speed and Scalability

FAISS enables efficient nearest-neighbor search, significantly reducing retrieval time.

Lower computational complexity (O(logN)) compared to traditional keyword search (O(N)).

Optimized for handling large-scale datasets, ensuring rapid policy retrieval.

Contextual Understanding and Accuracy

Captures semantic meaning beyond exact keyword matching.

Enables effective mapping between related terms (e.g., "oncology therapy" and "cancer treatment").

Improves policy selection accuracy by ensuring relevance to patient cases.

Storage Efficiency and Optimization

Embeddings condense large policy documents into high-dimensional vector representations.

FAISS allows for memory-efficient storage and retrieval, reducing overhead.

Dynamic and Adaptive AI Integration

Embeddings enable AI-driven retrieval without reliance on rigid, rule-based matching.

Enhances flexibility in handling diverse and evolving policy structures.

Supports retrieval-augmented generation (RAG) by feeding relevant policy data into GPT for decision support.