**Literature Review**

**1) Retrieval-Augmented Generation (RAG), embeddings, and ranking**

**RAG** combines a generator (LLM) with a non-parametric memory (document index). It consistently improves factuality and allows provenance, compared with generator-only systems. The seminal paper introduced RAG-sequence and RAG-token variants and showed state-of-the-art results on knowledge-intensive tasks. [arXiv+2NeurIPS Proceedings+2](https://arxiv.org/abs/2005.11401?utm_source=chatgpt.com)

For **dense retrieval**, Sentence-BERT (SBERT) gives efficient sentence embeddings for similarity search and clustering; it reduces pairwise BERT costs by orders of magnitude while preserving accuracy—practical for local FAISS indexes. [arXiv+2arXiv+2](https://arxiv.org/abs/1908.10084?utm_source=chatgpt.com)

**Vector search at scale** is productionized via **FAISS**, designed for high-dimensional approximate search with strong CPU/GPU performance and clear trade-offs across indexes (Flat, IVFFlat, HNSW, PQ). [arXiv+1](https://arxiv.org/abs/2401.08281?utm_source=chatgpt.com)

**Sparse retrieval** with **BM25** remains a strong lexical baseline; hybridizing dense+BM25 is robust in practice. [Wikipedia+1](https://en.wikipedia.org/wiki/Okapi_BM25?utm_source=chatgpt.com)

**Rank fusion**: **Reciprocal Rank Fusion (RRF)** is a simple, training-free method that almost always improves over the best individual ranker when combining multiple lists—ideal for your hybrid pipeline. [cormack.uwaterloo.ca+1](https://cormack.uwaterloo.ca/cormacksigir09-rrf.pdf?utm_source=chatgpt.com)

**Re-ranking**: Lightweight **MS MARCO Cross-Encoders (MiniLM-L6-v2)** provide strong passage re-ranking after first-stage retrieval. [Hugging Face+1](https://huggingface.co/cross-encoder/ms-marco-MiniLM-L6-v2?utm_source=chatgpt.com)

**2) Swim performance & training science (what the bot’s advice should align with)**

**Pacing/intensity estimation**

* **Critical Swim Speed (CSS)/Critical Velocity (CV)** — practical index of aerobic capacity; derived from distance-time linear modeling (e.g., 200/400 m tests) and widely used for threshold/tempo targeting. [traplago.com+3PubMed+3PubMed+3](https://pubmed.ncbi.nlm.nih.gov/1555562/?utm_source=chatgpt.com)

**Tapering**

* Taper (systematic reduction in load before competition) yields ~0.5–6% performance gains on average; responses depend on adjustments to volume, intensity, and frequency. [ScienceDirect+2PubMed+2](https://www.sciencedirect.com/science/article/abs/pii/S0765159711001213?utm_source=chatgpt.com)

**Injury epidemiology (shoulder-first)**

* Shoulder pain is highly prevalent across competitive swimmers; risk relates to volume, technique, and history. Recent systematic reviews and large surveys quantify high rates and encourage surveillance + targeted strength/mobility. [ijspt.scholasticahq.com+3PubMed+3PMC+3](https://pubmed.ncbi.nlm.nih.gov/37515375/?utm_source=chatgpt.com)

**Nutrition & recovery**

* Consensus statements (ACSM/Academy/DC) and IOC guidance underpin carbohydrate (~30–60 g·h⁻¹ for long sessions), protein timing, hydration, and safe supplement use; REDs (Relative Energy Deficiency in Sport) screening is emphasized. [Olympics+5PubMed+5drugfreesport.org.za+5](https://pubmed.ncbi.nlm.nih.gov/26891166/?utm_source=chatgpt.com)

**Records, rules, and governance**

* Official **World Aquatics** database (formerly FINA) is the source of record lists and rules; the federation rebranded in 2022/2023. Use their records endpoints/pages for authoritative facts. [World Aquatics+2World Aquatics+2](https://www.worldaquatics.com/swimming/records?utm_source=chatgpt.com)

**Monitoring load & readiness (optional signals)**

* **HRV** (e.g., rmSSD/Ln-rmSSD) can track adaptation and guide endurance training; best practices emphasize multiple weekly samples and caution about interpretation/context. [Thieme+3PubMed+3PubMed+3](https://pubmed.ncbi.nlm.nih.gov/23852425/?utm_source=chatgpt.com)

Takeaway: the chatbot should align its prescriptions with CSS/threshold practice, established tapering principles, shoulder-injury prevention basics, and mainstream sports-nutrition positions; for facts/records, cite World Aquatics.