

Systematically Improving RAG

Jason Liu

Playbook

- Nail the Feedback mechanism (Satisfaction)
- Monitor cosine & rerankers (Relevancy)
 - query, cosine, feedback
- Need this because not every query will have feedback
- Cluster queries to understand
 - topic, count, mean cosine, mean feedback
 - Rank importance of question types against relevance & satisfaction
 - Look at data to propose individual experiments to improve individual clusters
 - Build classifiers to improve obs in real time
- low hanging fruit (if we have time)
 - use synthetic data to generate baselines for search
 - Make sure to use BM25 + rerankers (where)
 - Include chunk metadata in a query-able way.
 - Include Date filters, Doc Type filters.