Leveraging Retrieval-Augmented Generation (RAG) and Knowledge Graphs to Enhance Lymphoma Clinical Research and Decision Support

Michael J. Kane and Christopher Flowers
MD Anderson Cancer Center
*Corresponding author: mjkane@mdanderson.org

Background

- Rapid growth of biomedical literature, genomics, and trial data challenges lymphoma research synthesis.
- Fragmented data across publications, guidelines, genomic and real-world datasets.
- AI combined with domain-specific knowledge offers new ways to generate hypotheses and guide decisions.
- AI is often most effective when it is **application oriented** – focusing on augmenting existing tasks.
    - Effectiveness can be quantified by how much time is saved for tasks.
    - Provides new opportunities from the automation of previously time-intensive tasks.

Objective

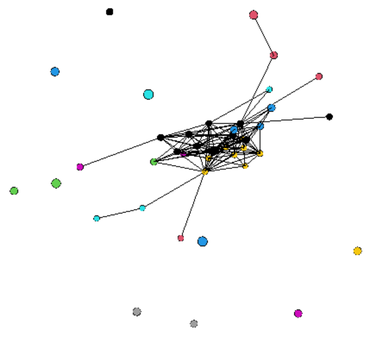
Develop **LYRA** (*Lymphoma Retrieval-Augmented Assistant*), an AI platform combining **RAG** and **knowledge graphs** to:

1. Retrieve & synthesize lymphoma research.
2. Support translational research and protocol development.
3. Identify high-impact opportunities for human augmentation and automation
4. Provide reproducible and validated results.

Architecture and Functionality

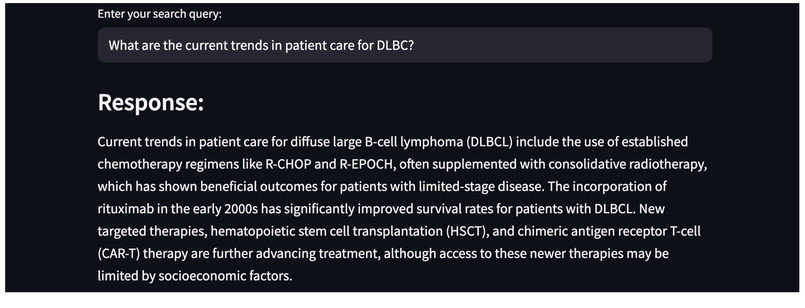
System Architecture

1. **Data Ingestion:** Publications, guidelines, registries, protocols
2. **Entity Extraction:** Named entities (disease, biomarkers, drugs, endpoints)
3. **Embedded Search:** Semantic representations for search & retrieval
4. **Knowledge Graph:** Integration across data sources
5. **RAG Interface:** Combines retrieval & generative AI for tailored responses
6. **Validation:** Ensures clinical & factual accuracy



Document Proximity

Black cluster – Prognostic Biomarker evaluation
Yellow cluster – IT efficacy and Safety
Red cluster – Real-world outcomes



Applications (in progress)

- **Novelty Quantification:** Detect emerging topics via similarity metrics
    - Biomarkers or altered metabolic pathways
    - Identify gene-function relationships
- **Expertise Mapping:** Author–topic networks for collaborator identification
    - Researcher Collaboration and Impact
    - Grant and Funding Matching
- **Therapeutic Trend Analysis:**
    - Drug repurposing
    - Propose novel drug/marker combinations
- **Predictive Modeling:** Structured trial variables extracted for forecasting.
    - Patient trial matching
    - Outcome prediction

Conclusion and Clinical Impact

RAG + knowledge graphs form a powerful framework for hematologic oncology research

LYRA demonstrates feasibility and scalability for intelligent lymphoma decision support

Impact

- Accelerates early-phase trial planning
- Supports strategic investigator collaboration and funding proposals
- Bridges translational and clinical lymphoma research