Proposal

AstroBio Knowledge Explorer

Executive Summary

NASA has conducted extensive biological experiments in space for decades, producing a vast corpus of research on microgravity biology, plant growth, human physiology, and microbial behavior in extraterrestrial environments. While publicly available, this knowledge is scattered across thousands of publications, making it difficult for scientists, engineers, and educators to locate information relevant to upcoming missions to the Moon and Mars.

Our project, *AstroBio Knowledge Explorer*, will build an AI-powered dynamic dashboard that leverages knowledge graphs, natural language processing (NLP), and advanced retrieval methods to summarize and visualize NASA bioscience publications. By structuring findings into an interactive, searchable interface, users can quickly explore key biological outcomes, compare experiment contexts, and validate information with direct links to original sources.

Problem Definition

NASA has produced decades of biological research on human health, plant growth, microorganisms, and cellular processes in microgravity. Yet this knowledge is scattered across thousands of unstructured reports and datasets with inconsistent terminology and limited metadata. Conventional keyword search cannot capture semantic meaning, link related findings, or filter by mission context, making it difficult for researchers to efficiently locate and apply relevant insights.

As space agencies prepare for lunar bases and Mars missions, rapid access to validated bioscience knowledge is critical. The main technical challenge is transforming heterogeneous, unstructured literature into a structured, queryable knowledge base. Without tools that extract entities, normalize terminology, and connect results into an integrated knowledge graph, NASA's bioscience legacy remains underutilized—slowing scientific discovery and hindering mission preparation.

Background & Literature Review

Existing efforts to manage scientific literature include digital repositories such as PubMed, NASA's Technical Reports Server (NTRS), and specialized archives like NASA Life Sciences Data Archive. These platforms provide access but rely heavily on keyword searches,

limiting their ability to support semantic exploration or pattern discovery. More advanced tools like Semantic Scholar and OpenAlex apply machine learning to scientific text, offering citation analysis and related paper discovery, but they are not tailored to the unique challenges of bioscience in space.

Recent advances in AI — particularly retrieval-augmented generation (RAG) and knowledge graphs — provide new opportunities for knowledge discovery. Knowledge graphs have been successfully applied in domains like biomedical research (e.g., Hetionet, Monarch Initiative) to integrate heterogeneous data and reveal hidden relationships. In parallel, dashboards that combine visualization with natural language summaries (e.g., Litmaps, Connected Papers) have shown how AI can help researchers quickly orient themselves within a literature space.

However, no existing system integrates NASA's bioscience corpus into a structured, dynamic knowledge graph accessible through an intuitive dashboard. Our project addresses this gap by combining AI-based literature summarization with graph-based exploration, specifically tailored to highlight experiment impacts and results relevant to long-term human space exploration.

Methodology

1. **Corpus Preparation**: Collect NASA bioscience publications from NTRS and related repositories.

2. **Text Processing**: Use NLP to extract entities (e.g., organisms, biological processes, mission contexts) and relations (e.g., "microgravity impacts bone density").

3. **Knowledge Graph Construction**: Organize extracted information into nodes and edges with provenance metadata.

4. **AI Summarization**: Apply large language models to generate concise experiment summaries validated against the knowledge graph.

5. **Dashboard Development**: Design an interactive UI for filtering (by organism, mission, biological system), visualizing experiment impacts, and comparing results.

6. **Integration**: Connect the knowledge engine with the UI to provide real-time query responses and graph-based exploration.

7. **Validation**: Ensure factual accuracy by aligning AI outputs with the structured knowledge graph and original publications.

Solution

Our solution tackles the challenge by combining structured knowledge representation with user-friendly exploration:

- The **knowledge engine** ingests bioscience publications, extracts structured insights, and builds a dynamic knowledge graph with provenance.
- The **dashboard** allows users to search, filter, and visualize results, enabling quick discovery of relevant experiments.
- AI-powered summaries and graph queries ensure that responses are accurate, relevant, and easy to understand.

To implement successfully, we need:

1. Data ingestion and preprocessing pipelines.
2. Knowledge graph database (Neo4j).
3. NLP/LLM-based entity-relation extraction and summarization.
4. UI/UX framework (React + D3.js for visualization).

This solution aligns with NASA's requirements by making bioscience findings more accessible, discoverable, and actionable for future mission planning.

Value Proposition

Our solution transforms NASA's vast but underutilized bioscience research into an accessible and interactive tool. For researchers, it reduces the time to find relevant studies. For mission planners, it consolidates biological risks and solutions. For educators, it provides engaging summaries and visualizations of space biology. Compared to static archives or keyword-based search engines, our AI+KG approach enables semantic discovery, highlights experiment impacts, and validates findings with provenance. This ensures higher accuracy and deeper insights than alternatives. By bridging raw literature and actionable knowledge, our solution maximizes the return on NASA's decades-long investment in bioscience research.

Role of Team Members

Zeyad Shahin: Responsible for building the NLP pipelines, knowledge graph, AI summarization, and integration with backend systems.

Mohammed AlMansori: Responsible for designing and developing the interactive dashboard, user workflows, and visualization tools to ensure accessibility and engagement.

Workflow Strategy

1. **Problem Analysis**: Review challenge requirements and data sources.

2. **Backend Development** (Zeyad): Build ingestion, extraction, and knowledge graph modules.

3. **Frontend Development** (Mohammed): Create wireframes, implement dashboard components, and link to backend.

4. **Integration & Testing**: Connect knowledge engine to dashboard; validate results with sample queries.

Resources

Software: Python (spaCy, Neo4j, Hugging Face transformers), React, D3.js, Docker

Hardware: Cloud GPU instances (for NLP/LLM)

Data: NASA Technical Reports Server (NTRS), Life Sciences Data Archive

Team Name: MCs
1- Zeyad Shahin
2- Mohammed AlMansori