**Project Progress Report**

Md Farukuzzaman  
COMP 4911

The University of Memphis

March 14, 2025

1. Introduction

This report outlines the progress made in enhancing the Merjek-AI Project, an AI-powered search engine designed to improve accessibility and engagement for university websites. The project focuses on refining search capabilities through AI-driven solutions, addressing key limitations of traditional search engines, and providing universities with a smarter, more intuitive search system.

2. Project Objectives

The primary objectives of this project are:

* Enhancing Merjek-AI to improve search accuracy and efficiency.
* Implementing natural language processing (NLP) for better user queries.
* Expanding support for multi-language searches to accommodate diverse user needs.
* Ensuring institutional control over search data while maintaining comprehensive search coverage for university content.
* Optimizing search results to understand user intent rather than relying solely on keyword matching.

3. Issues with Traditional Search Engines

Traditional search engines face several challenges that hinder accessibility and usability:

* Missing Links : Pages that are not directly linked on a university’s website are often unsearchable, leaving important resources hidden.
* Limited Natural Language Support : Users must rely on exact keywords rather than searching in everyday language.
* No Multi-Language Functionality : Most search engines restrict results to the primary language of the content, creating barriers for non-native speakers.
* Lack of Institutional Control : Universities depend on external search providers, reducing control over their search infrastructure.
* Keyword-Centric Results : Standard search engines prioritize keyword frequency rather than understanding context or intent behind queries.

4. Solution: The Merjek-AI Approach

To overcome these limitations, Merjek-AI introduces several innovative features:

* Comprehensive Coverage : The system ensures unlinked pages are indexed using customized AI-generated datasets.
* Natural Language Understanding : Users can enter queries in everyday language, and AI will interpret intent to return the most relevant results.
* Multi-Language Capabilities : The search engine is designed to support multiple languages with the appropriate prompt datasets.
* Institutional Data Ownership : Universities retain complete control over search data and indexing preferences.
* Context-Aware Results : AI-driven search leverages semantic understanding to deliver results that align with user intent rather than just keyword occurrence.

5. Progress to Date

Significant progress has been achieved, building upon the proof-of-concept study completed by the Fall 2024 capstone team and project owner, Dr. Fatih Șen.

Early Implementation:

* Developed and tested custom AI models for prompt generation and optimization.
* Experimented with large language models (LLMs) to improve search accuracy.
* Evaluated multi-GPU performance to enhance AI processing speeds.

Model Evaluations:

Several models have been assessed to determine their effectiveness in understanding search intent:

* GPT-based models (OpenAI, DeepSeek) for broad search capabilities.
* Fine-tuned smaller models for faster inference and better real-time search performance.
* Custom domain-trained models for university-specific search needs.

6. System Training & Performance Optimization

Cluster Configuration:

* Maximum Concurrent Jobs: 6
* Maximum Nodes: 3
* Maximum GPUs per Job: 4
* Maximum Runtime per Job: 48 hours

Training Progress:

* Dataset: 2,000 university webpages (subset of 10,000).
* Data Split: 80% training, 10% validation, 10% testing.
* Current Testing: Running on 2 GPUs (1 node).
* Training Time Estimate: 50 minutes per epoch.
* Projected Time for 10 Epochs: 8-9 hours.

The training phase has provided key insights into performance improvements, helping optimize AI models for real-time search applications.

7. Database Management & Migration

Initial Setup:

* The project started with a SQL database running on Windows, accessed via WSL Linux.
* A local MongoDB instance was used for preliminary data storage.

Migration to MongoDB Atlas:

* The system transitioned to MongoDB Atlas for cloud-based scalability and real-time query access.
* Integrated PostgREST for efficient data querying and retrieval.

Current Status:

* Local MongoDB setup completed.
* SQL to MongoDB migration successfully completed.
* Ongoing integration with MongoDB Atlas for cloud-based AI search indexing.

8. Next Steps & Future Plans

Short-Term Goals:

* Complete 10-epoch training on the 2,000-document dataset.
* Analyze and optimize GPU resource usage to enhance search efficiency.
* Scale AI model training from 2 GPUs to 4 GPUs for better performance.

Long-Term Goals:

* Expand search indexing to the full 10,000-document dataset.
* Optimize model accuracy using hyperparameter tuning.
* Ensure a seamless integration between AI-based search and university website data.

9. Conclusion

The project has made substantial progress in enhancing Merjek-AI as an intelligent search engine for university websites. The proof-of-concept model has been expanded with improved NLP capabilities, multi-language support, and institution-specific indexing control.

With ongoing optimizations, the next phase will focus on scaling the dataset, improving training efficiency, and ensuring seamless search integration for universities seeking better digital accessibility and engagement.