MISM Capstone Project Proposal

Project Title	**Buddy AI:** **Enhancing Employee Productivity and Engagement via a Private Enterprise LLM with Success Monitoring Capabilities.**

Project Sponsor
Hitachi America Ltd.

MSIM Project Team	
Name	**Role**
Haren Bhatia	Artificial Intelligence / Machine Learning Engineer
Devarsh Patel	Data Engineer and Analyst
Ananya Nandi	Product Manager

Project Overview

Project Summary

This project aims to develop Buddy AI, an intelligent support tool designed to improve employee productivity and engagement by providing personalized, efficient assistance. The Buddy AI system will be built using a multimodal Large Language Model (LLM) integrated with a knowledge graph to offer tailored answers to HR-related and other workplace queries. The project will focus on creating a secure, scalable AI infrastructure and data pipeline, supported by an interactive dashboard to measure usage, engagement, and BuddyAI's impact.

Goals and Objectives

The primary goal of the project is to develop Buddy AI, a secure, on-premise conversational agent designed to:

1. Assist employees with queries related to HR documents, policies, and other workplace information.
2. Provide real-time, personalized responses based on a rich knowledge base.
3. Integrate a data pipeline that collects, processes, and transforms SharePoint documents into a structured, usable format (golden state of data) for Buddy AI to query.
4. Visualize system usage, user engagement, and AI performance through an interactive dashboard to enhance stakeholder decision-making.

Key deliverables will include:

1. A working Buddy AI chatbot capable of answering employee queries.
2. A fully developed data pipeline that ingests, processes, and organizes HR-related documents (e.g., text, tables, images) from SharePoint into a structured Data Lake. The LLM will query this data.
3. An Analytics Dashboard to visualize usage metrics and engagement insights of Buddy AI

Key Deliverables

1. **Buddy AI Prototype**: A conversational AI system running on-premise that can efficiently handle employee queries using multimodal LLMs integrated with knowledge graphs.
2. **Data Pipeline**: A functional and secure data pipeline that collects HR documents (including tables, text, and images) from SharePoint and processes them into a golden state of data.

3. **Analytics Dashboard**: A tool (e.g., Grafana) to visualize and track system usage, response effectiveness, and overall employee engagement with Buddy AI.

Dependencies and/or related projects

1. SharePoint document integration (daily/weekly ingestion)
2. Kubernetes for on-premise hosting
3. Graph database for knowledge graph integration
4. Open-source LLMs fine-tuning frameworks (for customization)
5. Analytics dashboard integration (Grafana)

Technology and platform to be used

1. **Databricks Data Lake**: Secure data storage and processing solutions for HR documents.
2. **Kubernetes**: For scalable on-premise hosting of Buddy AI.
3. **Private Enterprise LLM**: Multimodal models for AI-based employee assistance with access to on prem servers.
4. **Graph Knowledge**: A graph database for structuring and querying knowledge.
5. **Grafana**: Dashboarding tools for system insights and analytics.
6. **Private AI Development Tools**: For secure, enterprise-grade AI model deployment.

Documentation Required

1. System architecture
2. Data pipeline design and implementation
3. Performance metrics and usage reports
4. Analytics dashboard design and insights
5. Deployment guide

Stakeholders

Name	Role	Contact Information
Krishnapillai, Bala	Vice President and Head of IT Group - Americas	Bala.Krishnapillai@hal.hitachi.com

High-level Timeline/Milestones

Milestone Description	Milestone Date
Project Initiation and Planning	01/30/2025
Technical stack research and alignments	02/15/2025
Data Pipeline Design and Integration	03/01/2025
Serving Buddy AI MVP on a private instance	03/30/2025
Analytics Dashboard Implementation	05/15/2025
Handing Over - Documentation and Presentation	05/30/2025

Success Metrics

- **Employee Engagement**: High usage rates and active interactions with Buddy AI, demonstrating its effectiveness in addressing HR-related queries.
- **Data Accuracy**: The AI's responses must be highly accurate, especially when addressing specific employee queries, with over 90% accuracy for structured documents.
- **Dashboard Insights**: A fully functional dashboard that provides actionable insights into system usage, engagement levels, and AI performance metrics.
- **Scalability**: The system should be scalable, with seamless deployment across multiple departments or teams within the enterprise.
- **Operational Efficiency**: The data pipeline must be optimized to transform and organize data efficiently, with minimal latency in query responses.
- **Compliance**: Ensuring all data storage, handling, and processing comply with relevant data privacy regulations (GDPR, etc.) and internal policies.

Risk Assessment

1. **Performance Constraints on Edge Hardware**

   The AI appliance may struggle to run fine-tuned LLMs efficiently due to limited computational resources on the edge device.
   Mitigation: Optimize model architecture (e.g., quantization, pruning), Use smaller, domain-specific LLMs rather than general-purpose models.

2. **Data Privacy and Security Breaches**

   Mishandling sensitive HR data during development or system breaches could compromise the project's objectives.
   Mitigation: Utilize synthetic or anonymized datasets during development and ensure no enterprise data is shared off-premise

3. **Regulatory Non-Compliance**

   The system may not meet all compliance requirements (GDPR - for analytics), especially for data storage and privacy.
   Mitigation: Consult legal and compliance experts during design.

4. **Stakeholder Misalignment and Scope Creep:**

   Time and resources may be constrained for a finetuned model, and a possible misalignment with stakeholders can also be possible in large projects
   Mitigation: Clear Documentation and Clear communication lines must be established as well as stakeholder reviews and thorough UAT Testing. Use Prototyping to ensure the progress is well tracked and visible to stakeholders.

5. **Data Pipeline Failures**

   Issues in the data pipeline could delay or impede the AI's ability to retrieve and process documents correctly.

   Mitigation: Implement strong monitoring, logging, and fallback mechanisms to ensure smooth data operations.

6. **Failure to Achieve Scalability**

   The prototype might not demonstrate scalability in enterprise-level environments.
   Mitigation: Focus on Scalable architecture that enterprises can plug and play with.

7. **Cost Overruns**

   Development costs (e.g., hardware run, and data preparation) may exceed the allocated budget.
   Mitigation: Define clear budgets, monitor expenses closely, and maintain transparency while frugally using tools

Assumptions

1. **Hardware Capabilities:**

   Hitachi's hardware will have sufficient computational power to run fine-tuned Large Language Models (LLMs) efficiently within the constraints of edge computing.

2. **Data Availability:**

   SharePoint documents will be consistently available for ingestion, and the data transformation process will run smoothly.

3. **Dashboard and User Interface:**

   Stakeholders will provide input and feedback to ensure the design meets industry standards.

4. **Stakeholder Support:**

   Stakeholders at Hitachi will be available for interviews and consultations to validate requirements and refine the use case. Clear and consistent communication channels

5. **Timeframe Feasibility:**

   The project timeline will be sufficient to complete hardware/software integration, fine-tuning, and stakeholder engagement within the proposed schedule. Delays due to unforeseen challenges will not significantly impact project deliverables.

6. **Scalability Requirements:**

   The prototype's scalability needs can be adequately simulated. If necessary, real-world scalability tests can be deferred to post-project phases.

7. **Focus on Privacy and Compliance:**

   Privacy and compliance requirements (e.g., GDPR) can be addressed using encryption, anonymization, and secure on-premises processing. The regulatory landscape will not change significantly during the project.

Sponsor Review		
Sponsor Name	**Position**	**Date**
Krishnapillai, Bala	Vice President and Head of IT Group - Americas	