Design and develop a versatile Multi-Modal Question Answering System – Project Plan

# Week 1: Introduction to Multi-modal Systems

## Tasks:

* Learn about question-answering systems and their main parts.
* Understand what embeddings are and why they matter in AI.
* Explore the basics of Multi-Modal AI systems.
* Explore existing multi-modal systems, understanding their architectures, strengths, and typical use cases.
* Explore the fundamentals of multi-modal question-answering systems.
* Familiarize yourself with the different kinds of multi-modal data that can originate in the manufacturing industry

## Objectives:

* Give the group a clear start on multi-modal systems, their use in answering questions, and their role in the manufacturing sector.

# Week 2: Deep Dive into Multi-modal Systems

## Tasks**:**

### Research and Design Group:

* Understand and document various Multi-Modal Systems and pinpoint the most dependable approaches to designing one.
* Clearly identify the parts and define the architecture of at least 3 different multi-modal system works.
* Basic research into fusion models and transformers, and how they fit in the context of multi-modal AI models.
* Define the main challenges and advantages of handling text, images, and tabular data at once.

### Coding Group:

* Get to know tools like Hugging Face, TensorFlow, and PyTorch for multi-modal models and tasks.
* Identify what's ready-made and what needs customizing in these libraries or tools offering multi-modal AI systems.
* Break down the parts of current systems: training, tuning, testing, customization, and deployment.
* Explore existing real-world applications that use Multi-Modal Systems.

## Objectives:

* Equip the team with a thorough understanding of multi-modal systems, their tools, and their practical applications.

# Week 3: Consolidation and Alignment

## Tasks:

* Both groups collaboratively:
  + Discuss and consolidate the information from Weeks 1 and 2.
  + Formulate key questions and potential challenges related to developing a multi-modal system.
  + Research and Design group outlines potential techniques and methods worth pursuing.
  + The coding group proposes possible implementation routes and tools/frameworks to be used.
  + Both teams must together define the building blocks for a custom pipeline we aim to achieve in the remaining weeks.

## Objectives:

* Synchronize knowledge and insights of both groups.
* Define a clear roadmap for design and coding in the upcoming weeks.

# Week 4 & 5: Design and Prototyping of the Multi-modal System

## Tasks in Research and Design:

* Break down and define components for multimodal systems. Highlight how different modalities (text, images, table data) will be integrated into the systems.
* Design a potential architecture for a multi-modal question-answering system
* Highlight limitations of the system and also areas that need specific improvement once the system is made modular.
* What method of pre-fusion and post-fusion methods to choose and what are the trade-offs to be considered for speed, and accuracy?

## Tasks in Coding:

* Establish a local Python environment and a cloud instance to run; initiate key ML/DL tools and libraries.
* Prototyping:
  + Deploy two distinct online multi-modal systems for testing.
  + Investigate existing multi-modal datasets; code to identify/ correlate key features that make such datasets work for multi-modal systems.
  + Assess a general-purpose multi-modal Q&A system.
  + Ensure experiments with models are backed by functional API and SDK coding. Not just GUI setup.
  + Analyze the potential of model fine-tuning, giving preference to simpler, efficient models (e.g., from Hugging Face).

## Objectives:

* Convert research into practical designs.
* Begin the development and prototyping of the system, ensuring alignment with design proposals.

# Week 6 & 7: System Enhancement and Testing

## Tasks in Research and Design:

* Determine necessary data transformations and metadata needed to standardize and enrich datasets for multi-modal tasks.
* Identify evaluation metrics for gauging multi-modal system performance.
* Represent a design for an architecture for both a fusion-based multi-modal Q&A system and fine-tuning a multi-modal Q&A system.
* Evaluate the feasibility of employing a fine-tuned large model for single modal systems, and further integrating results to enable Q&A.
* Study emerging architectures in multi-modal domains that promise enhanced Q&A capabilities while maintaining modularity for dataset-specific adjustments.

## Tasks in Coding:

* Engage in testing using the company’s custom dataset.
* Define and code the chosen evaluation metrics for result validation.
* Design modules to preprocess the multi-modal dataset aimed at fine-tuning.
* Based on our fusion strategy, either:
  + Craft a fusion pipeline for model inference, or
  + Identify and structure the fine-tuning code for model adjustments with custom datasets.
* Construct pipeline code with APIs to:
  + Designate the path for multi-modal data during fine-tuning.
  + Manage versioning of fine-tuned models
  + Extract Q&A results, be it from a single-modal or multi-modal system, or using a tailored wrapper for a fine-tuned model.

## Objectives:

* Enhance and fine-tune the multi-modal system.
* Guarantee consistent and reliable performance for diverse data sources and queries.

# Week 8: Summarization, Reporting, and Feedback

## Tasks:

* Select content for university presentation, keeping privacy and ownership in mind.
* Finalize the codebase and design documents.
* Summarize project highlights, breakthroughs, challenges, and potential next steps.
* Engage with the Braviz team to receive feedback on the system and its potential integration into existing products.

## Objectives:

* Get the project ready for university presentation.
* Understand key outcomes and formulate a plan for potential future developments or improvements.

# General Notes:

* Design tasks must be visually represented using a digital whiteboarding tool provided by the company, aiming to address previously formulated questions.
* All coding is in Python, use GUI-based setups only for experiments. Ensure regular commits to the Git repository provided.

# Early Reading Suggestions:

* <https://www.kdnuggets.com/2023/03/multimodal-models-explained.html>
* <https://huyenchip.com/2023/10/10/multimodal.html>