SSY226 - Planning Report

group number 15

Abhijeet Singh Dhillon (dhillon)
Mirarash Keshavarz Kelachayeh (mirkes)
Sanam Molaee (molaee)
Daniel Söderqvist (danisode)

December 12, 2023

Title

Design and Develop a Versatile Multi-Modal Question Answering System

Background

Since humans perceive and understand the surrounding environment through various kinds of data such as text, images and sounds, one of the approaches in artificial intelligence is to mimic human behaviour using multi-modal models. Nowadays, there has been increasing interest in Question Answering (QA) models that works based on multiple datasets. The idea behind this project is to design a multi-modal Question Answering system that is able to process textual and visual data from various sources for the industry. While the use of AI and Machine Learning is becoming popular in every industrial field, Braviz aims to focus on translating the complex industrial data as user-friendly as Google.

Purpose

The purpose of this project is developing a Question Answering system through investigation of the basics of multi-modal systems in terms of structure, performance of each specific method, and implementation of at least two multi-modal models such as CLIP, Flamingo and so on in pytorch library. To do so, several architecture of multi-modal systems will be studied and then their structure will be investigated. The next step will be the integration of different modalities, including text, image, and tabular data, to the model. The best feasible method will be selected based on trade off accuracy, solution time and other metric evaluations. At the end, an architecture for the multi-modal Question Answering system and fine tuned model will be proposed.

Problem/Task

Through this project the following questions associated with Multi-Modal Question Answering System will be addressed:

- Fundamentals of Modalities: Different kinds of modalities will be studied in terms of architecture, performance, pros and cons, limitations and so on.

- Representation of Modalities: It deals with how various types of modalities can be represented for machine learning processes.
- Interconnection of Modalities: The interaction and connection of the different types of the mentioned data representations will be done.
- Investigation on Open-Source Multi-modal Systems: In this section, the aim will be investigation of practical aspects such as finding prebuilt multi-modal models, how their architecture is defined, which loss functions which will be used and so on.
- Challenges: At this stage, the assessment of challenges and limitations of building multi modal systems will be addressed.
- Dataset: This stage deals with addressing some questions on datasets i.e if there are any available datasets for industry, if not how a dataset can be built.
- Conclusion: At the end, all the above tasks will be summarised to develop a multi-modal Question Answering system.

Boundaries

One of the main boundaries and concerns are the completion of the task and feasibility of the practical implementation within the given study period. This is due to the fact that Machine Learning applications may be time consuming.

Method/Implementation

The methodology of this project consists of multiple parts, including:

- Research on the fundamental of multi-modal systems.
- Research on different model architecture and techniques in order to design the models as well as possible implementation methods.
- Select two potential architectures for the multi-modal Question Answering system.
- Deploy two distinct online multi-modal systems for testing.

- Engage in testing and validation using the company's custom dataset.
- Construct pipeline codes with APIs to extract QA results.

Timetable

For time management and main structure of the planning, a gantt-chart will be used. This will help keep track of progress and important deadlines. The different parts in the gantt-chart are mainly divided into weekly tasks that are dependent on previous tasks. Writing the report for the project will be a parallel task that depends on the progression of the project. The gantt-chart is shown in the figure below.

Design and Develop a Versatile Multi-Modal Question Answering System

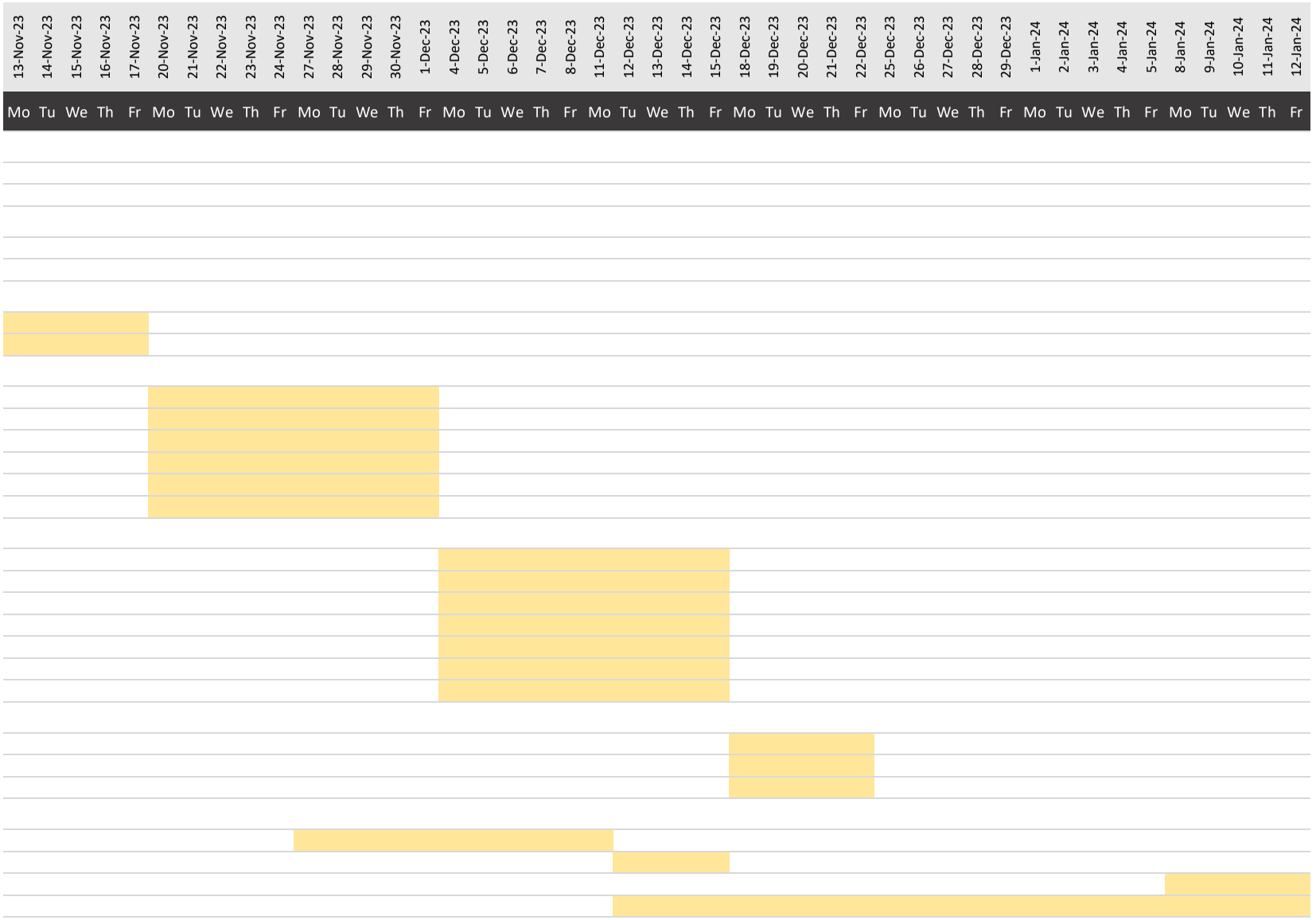
Project leader: Sanam
Project start date: 30 oktober 2023

Week: 1

☑ Show to late
☐ Show weekend
☐ Show member colors

TASKS	RESPONSIBLE	PROGRESS	WORKING DAYS	START	FINISH
Introduction to Multi-modal Systems					
Learn about Multi-modal systems	**All**	80%	5	30-Oct-23	5-Nov-23
Planning report	**All**	60%	5	30-Oct-23	5-Nov-23
Deep Dive into MM Systems					
Thorough understanding of arcitecture and models	**Research group**	0%	5	6-Nov-23	12-Nov-23
Thorough understanding of libraries and coding tools	**Code group**	0%	5	6-Nov-23	12-Nov-23
Consolidation and Alignment					
Sync knowledge between groups	**All**	0%	5	13-Nov-23	19-Nov-23
Plan clear roadmap for upcoming week	**All**	0%	5	13-Nov-23	19-Nov-23
Design and Prototyping of the MM System					
Break down components of MM systems	**Research group**	0%	10	20-Nov-23	3-Dec-23
Design potential arcitecture for MM questioning system	**Research group**	0%	10	20-Nov-23	3-Dec-23
Investigate limitations and pre vs post-fusion	**Research group**	0%	10	20-Nov-23	3-Dec-23
Establish library and cloud instance	**Code group**	0%	10	20-Nov-23	3-Dec-23
Deploy 2 online MM systems	**Code group**	0%	10	20-Nov-23	3-Dec-23
Investigate existing models and general-purpose system	**Code group**	0%	10	20-Nov-23	3-Dec-23
System Enhancement and Testing					
Data standardization	**Research group**	0%	10	4-Dec-23	17-Dec-23
Performance evaluation	**Research group**	0%	10	4-Dec-23	17-Dec-23
Arcitecture design	**Research group**	0%	10	4-Dec-23	17-Dec-23
Integration and exploration	**Research group**	0%	10	4-Dec-23	17-Dec-23
Testing and dataset preparation	**Code group**	0%	10	4-Dec-23	17-Dec-23
Data preprocessing and fusion strategy	**Code group**	0%	10	4-Dec-23	17-Dec-23
Pipeline development and management	**Code group**	0%	10	4-Dec-23	17-Dec-23
Summarization, Reporting and Feedback					
Prepare presentation for Braviz	**All**	0%	5	18-Dec-23	24-Dec-23
Finalize codebase	**All**	0%	5	18-Dec-23	24-Dec-23
Summarization of project and engaging with Braviz	**All**	0%	5	18-Dec-23	24-Dec-23
Report					
Draft report	**All**	0%	11	27-Nov-23	11-Dec-23
Peer review of draft report	**All**	0%	4	12-Dec-23	15-Dec-23
Presentation	**All**	0%	5	8-Jan-24	12-Jan-24
Final report	**All**	0%	24	12-Dec-23	12-Jan-24

Settings


13-Nov-23
14-Nov-23
15-Nov-23
16-Nov-23
17-Nov-23
20-Nov-23
21-Nov-23
22-Nov-23
23-Nov-23
24-Nov-23
27-Nov-23
28-Nov-23
29-Nov-23
30-Nov-23
1-Dec-23
4-Dec-23
5-Dec-23
6-Dec-23
7-Dec-23
8-Dec-23
11-Dec-23
12-Dec-23
13-Dec-23
14-Dec-23
15-Dec-23
18-Dec-23
19-Dec-23
20-Dec-23
21-Dec-23
22-Dec-23
25-Dec-23
26-Dec-23
27-Dec-23
28-Dec-23
29-Dec-23
1-Jan-24
2-Jan-24
3-Jan-24
4-Jan-24
5-Jan-24
8-Jan-24
9-Jan-24
10-Jan-24
11-Jan-24
12-Jan-24
Mo Tu We Th Fr Mo Tu We Th Fr Mo Tu We Th Fr Mo Tu We Th Fr Mo Tu We Th Fr Mo Tu We Th Fr Mo Tu We Th Fr Mo Tu We Th Fr Mo Tu We Th Fr