School of Electronic Engineering and Computer Science
Queen Mary University of London

MSc PROJECT DEFINTION
2022-23

This project definition must be undertaken in consultation with your supervisor. The feasibility of the project should have been assessed and the project aims should be clearly defined.
Submission of this document implies that you have discussed the specification with your supervisor.

Project Title: Controlled Unsupervised Multimodal Machine Translation using Visual Information

Supervisor: Julia Ive

Student name: Gargeya Sharma

Student e-mail: ec22146@qmul.ac.uk

PROJECT AIMS:

The project aims to tackle the problem of unsupervised (controlled) multimodal machine translation while working with visual information as the secondary mode of information to guide the learning process. Until now, multimodal machine translation problems were solved in a supervised manner, so this project aims to extend the application without direct human intervention (supervised learning). This approach will be using reinforcement learning with pre-trained language models to solve its purpose.

PROJECT OBJECTIVES:

The main project objective is to use the skills and added knowledge attained from the ongoing master's program and introduce novelty in the ongoing A.I. research with the effort to contribute towards a more powerful and safe ecosystem involving intelligent agents among us.

My course stream is aligned towards computer vision and working especially with visual data. So, while utilizing that knowledge, the project aims to incorporate visual information (images) as the guiding force towards performing the task of machine translation and improving the accuracy and fluency of those translations, making them more useful for a wide range of applications. The use of multimodal information, such as

both text and images in our case, is an example of how artificial intelligence can integrate and make sense of data from multiple sources.

This approach to machine translation also involves the use of unsupervised learning, which is a type of machine learning that relies on large amounts of unannotated data to learn patterns and structures in that data to perform the required task.

From the implementation point of view, one of the objectives is to define the problem in terms of reinforcement learning problem where the reward function is based on visual information.

All these objectives aim to prove my competency in developing artificial intelligence solutions with different modes of data and being able to identify and solve research gaps in the domain.

METHODOLOGY:

1. *Data collection and preparation*: Collect/Find a large corpus of parallel text in the source and target languages along with corresponding images and preprocess them to clean and normalize them.
2. *Multimodal representation*: Design the discrete multimodal space to learn the relationship between different modes of data used in this project.
3. *Pre-trained model*: Ready and fine-tune a pre-trained language model to act as the base models for agents used in reinforcement learning to achieve the task.
4. *Formulize the problem as an architecture*: Using all the necessary blocks attained so far, problem formulation in the form of reinforcement learning's environment and reward function defined using visual information needs to be carried out so that the experimentations with the approach can begin.
5. *Experimentation:* Iteratively experiment with multiple approaches for each of the steps to produce the most satisfactory results using multiple evaluation metrics for comparative analysis.
6. *Finalize:* After thorough research and reliable, repeatable results we will try to achieve the right combination of a properly functional pipeline that states that the aims of the project are fulfilled.

PROJECT MILESTONES

- Curated multimodal dataset to solve tasks related to the problem domain.
- Working software in the form of python notebooks or python files to represent the fulfilment of this project's aim while also ensuring repeatability and open-source sharing of the implementation to be used in further research on the topic.

- Detailed report in the form of a document comprising every detail about the failures and successes experienced throughout the journey of this project's fulfilment.

REQUIRED KNOWLEDGE/ SKILLS/TOOLS/RESOURCES:

- Programming skills with Python language
- Fluent use of PyTorch framework to develop and train the resultant system.
- Sufficient understanding of applied reinforcement learning, machine translation, multimodal representations and visual data characteristics to define a suitable reward function for the system.
- GPU resources to train the multimodal approach using reinforcement learning.
- Research skills to understand and select the literature on this topic and modify any useful open-source software to fit certain purposes for the final implementation.
- Data curation techniques to properly align the data with the problem statement.

TIMEPLAN

10th May:
Start the project and read a bunch of literature to formulate a well-defined pathway and a rough architecture of the solution as a proof of concept.

1st June:
Start curating the dataset according to the need of the architecture. Parallelly, research open-source software which can be used as some of the building blocks in my implementation and try to reproduce their outputs. If nothing works out, writing out the code for everything I need from scratch.

25th June:
A working code as a proof of concept for further experimentation is ready to be used. Start recording the process of experimentation and modification in a document which further will be used as a reference point to finalize the draft of the dissertation research paper to be submitted before 10th July.

15th July:
Enough experimentation has been carried out to reach a conclusive design of the solution. More refinement in the code and approach might still be active to achieve repeatability, reliability, readability, and reusability of the code. Parallelly, we will start creating a draft for the dissertation report and maintain a list of requirements such as performance results, diagrams etc. to be included in the report.

15th August:

The working software is ready and complete while the dissertation report is with the supervisor to be assessed for one final review before submission.

~22nd *August:*

End the project by submitting all the required documents for the evaluation.