

# Configuration Manual

MSc Research Project  
MSCAI1-Master of Science in Artificial Intelligence

**Anu Benny**  
StudentID: x23104937

School of Computing  
National College of Ireland

Supervisor: Prof. Mayank Jain

**National College of Ireland  
Project Submission Sheet  
School of Computing**

<b>Student Name:</b>	Anu Benny	
<b>Student ID:</b>	x23104937	
<b>Programme:</b>	MSCAI1 - Master of Science in Artificial Intelligence	
<b>Year:</b>	2024	
<b>Module:</b>	MSc Research Project	
<b>Supervisor:</b>	Prof. Mayank Jain	
<b>Submission Due Date:</b>	12/08/2024	
<b>Project Title:</b>	Configuration Manual	
<b>Word Count:</b>	XXX	
<b>Page Count:</b>	5	

logos/NCI\_Logo\_colour.jpg

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

<b>Office Use Only</b>	
Signature:	

Date:	
Penalty Applied (if applicable):	

**ALL** internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	
Date:	12th August 2024

**PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:**

Attach a completed copy of this sheet to each project (including multiple copies).	<input type="checkbox"/>
<b>Attach a Moodle submission receipt of the online project submission</b> , to each project (including multiple copies).	<input type="checkbox"/>
<b>You must ensure that you retain a HARD COPY of the project</b> , both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	<input type="checkbox"/>

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

# Configuration Manual

Anu Benny  
x23104937

## 1 System Configuration

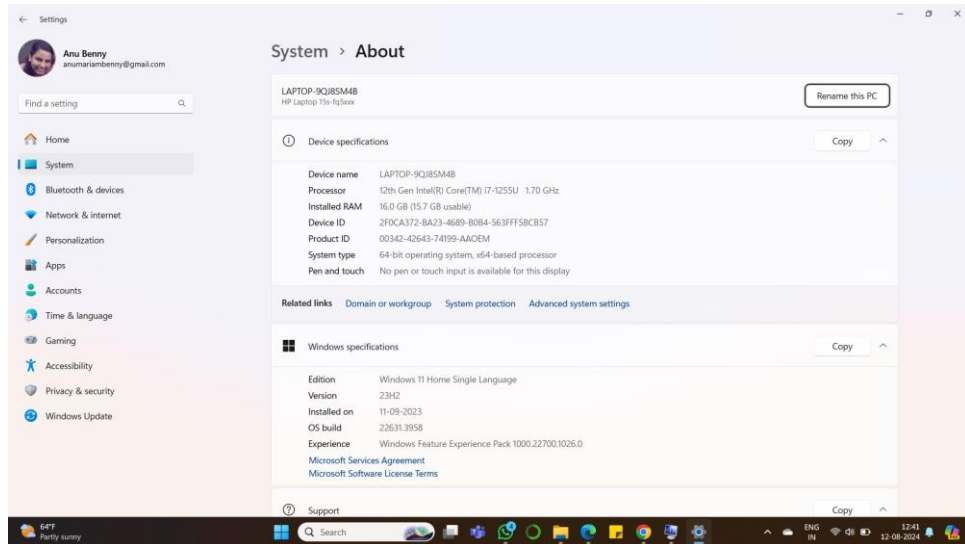


Figure 1: System configuration

## 2 Software Requirements

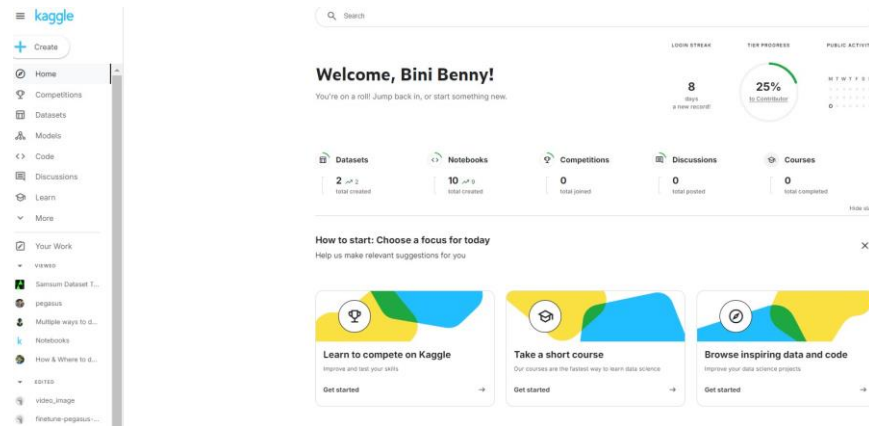
For the project we have used following software :

1. Kaggle Notebook
2. Kaggle GPU

## 3 Python Libraries

The project uses following python libraries :

1. Transformers
2. Pipeline
3. numpy
4. matplotlib
5. pandas
6. sklearn



7. pytesseract
8. opencv
9. nltk
10. tokenizer
11. YouTubeTranscriptApi
12. rouge score
13. Datasets

## 4 Dataset

The hugging face samsun dataset is used for fine tuning the pegasus model. The SAMSum dataset contains about 16k messenger-like conversations with summaries. **Data Fields** dialogue: text of dialogue. summary: human written summary of the dialogue. id: unique id of an example. **Data Splits** train: 14732 val: 818 test: 819

## 5 Finetuning the Pegasus

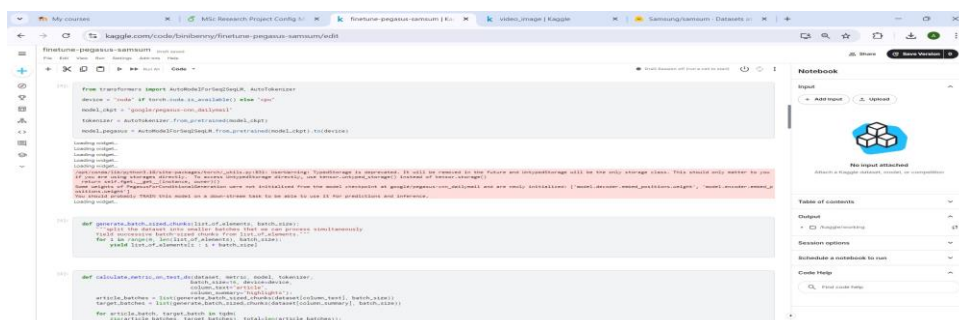


Figure 3: Preprocessing

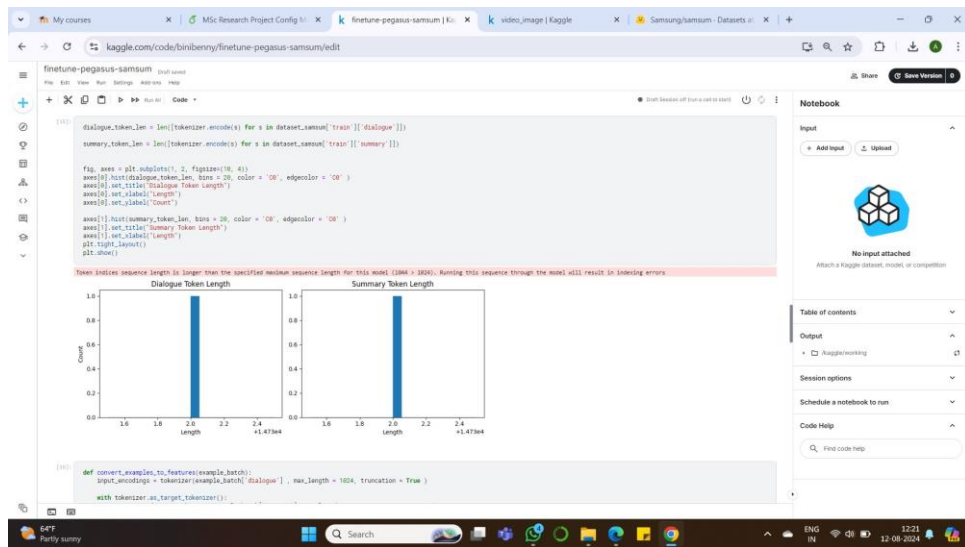


Figure 4: Histogram

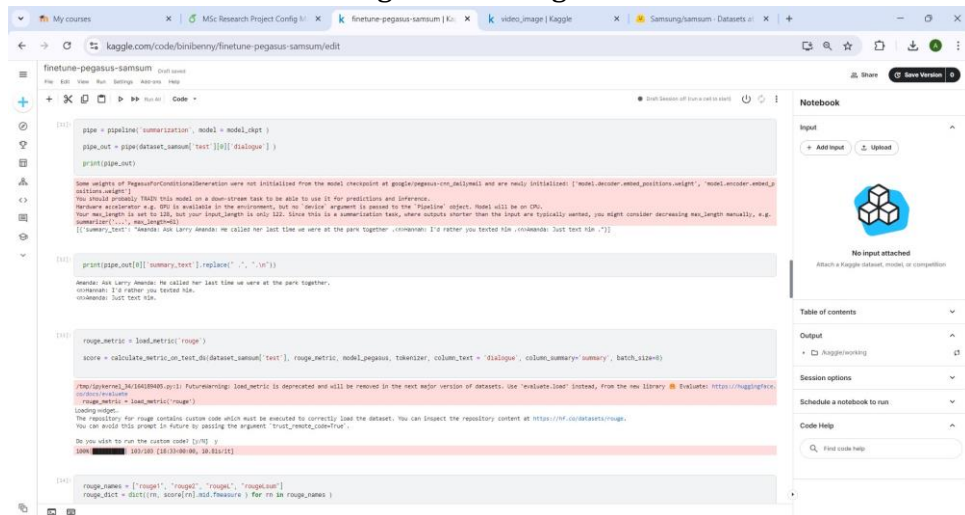


Figure 5: pipeline

