

Configuration Manual

MSc Research Project
MSc Artificial Intelligence

Mohammed Sameer Ahmed
Student ID: X23337257

School of Computing
National College of Ireland

Supervisor: Mr. Taylou Maniganze

National College of Ireland
MSc Project Submission Sheet
School of Computing



Student Name: Mohammed Sameer Ahmed
Student ID: X23337257
Programme: MSc Artificial Intelligence **Year:** 2024-25
Module: MSc Research in Computing
Lecturer: Mr. Taylou Maniganze
Submission Due Date: 13th September 2025
Project Title: Evaluating Deep Learning Models for Cross-domain Propaganda Detection
Word Count: 348 **Page Count:** 3

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.

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: Mohammed Sameer Ahmed

Date: 11th August 2025

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

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

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

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Configuration Manual

Mohammed Sameer Ahmed
Student ID: X23337257

1 System Requirements

This section provides the system requirements which is the first step in the process of reproducing the project files on a different machine. As this project was created and executed on a local system the system requirements are:

1. Modern quad-core CPU
2. 16GB RAM
3. 16GB GPU RAM
4. 35GB Storage
5. Python programming language

2 Data Resources

The data resources which are the project datasets and GloVe 100d file are already available in the zipped artefact. These resources are available from:

1. Propaganda Dataset (Wang *et al.*, 2020):
<https://github.com/leereak/propaganda-detection>
2. GloVe 100d file (Pennington, Socher and Manning, 2014):
<https://nlp.stanford.edu/projects/glove/>

3 Project Artefact

This section explains the project structure, figure 1 has the view of the project directory:

1. “data” folder contains the propaganda dataset used in the project in raw form.
2. “glove.6B” has the GloVe 100d file used in LSTM learning.
3. “lstm.py”, “bert.py”, “xlnet.py” and “gpt2.py” are the python scripts for the model.
4. “preprocessing.py” script runs to generate the cleaned dataset “clna.csv”, “clns.csv”, “clsa.csv”, “clss.csv”, “cltw.csv”.
5. “requirements.txt” will be used to install dependencies.
6. “charts.py” generates the graphs for the study.

Name	Type	Size
data	File folder	
glove.6B	File folder	
bert	Python Source File	5 KB
charts	Python Source File	3 KB
clna	Comma Separated Values Source File	34,994 KB
clns	Comma Separated Values Source File	1,072 KB
clsa	Comma Separated Values Source File	5,668 KB
clss	Comma Separated Values Source File	2,693 KB
cltw	Comma Separated Values Source File	1,316 KB
gpt2	Python Source File	5 KB
lstm	Python Source File	6 KB
preprocessing	Python Source File	2 KB
requirements	Text Document	4 KB
xlnet	Python Source File	5 KB

Figure 1: Project Directory

4 Environment Setup

To run the project files the following steps need to be done in order:

1. Extract artefact file and create a python virtual environment in the directory:
`python -m venv venv`
2. Activate the virtual environment:
`venv\scripts\activate`
3. Install the dependencies using “requirements.txt”:
`pip install -r requirements.txt`

5 Execution Steps

To execute the project files the following steps need to be done in order:

1. Run “preprocessing.py” to clean the raw data into dataset “clna.csv”, “clns.csv”, “clsa.csv”, “clss.csv” and “cltw.csv”.
2. Run the models “lstm.py”, “bert.py”, “xlnet.py” and “gpt2.py” to see their performance.
3. To change the domain change the file name at training and comment that file at inference. For example, “clns.csv” is used for training and commented at inference in the code snippet in figure 2.

```

...      ...

9      model_save_dir = 'bert_results/ns'
10     train_data_path = 'clns.csv' # Change this to your training data path
...      ...

122    # Comment out the training dataset below
123    inference_bert('clna.csv')
124    # inference_bert('clns.csv')
125    inference_bert('clsa.csv')
126    inference_bert('clss.csv')
127    inference_bert('cltw.csv')

```

Figure 2: Code Snippet

4. Run the “charts.py” to view the bar graphs of performances.

References

Pennington, J., Socher, R. and Manning, C. (2014) ‘GloVe: Global Vectors for Word Representation’, in A. Moschitti, B. Pang, and W. Daelemans (eds) *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. *EMNLP 2014*, Doha, Qatar: Association for Computational Linguistics, pp. 1532–1543. Available at: <https://doi.org/10.3115/v1/D14-1162>.

Wang, L. *et al.* (2020) ‘Cross-Domain Learning for Classifying Propaganda in Online Contents’. arXiv. Available at: <https://doi.org/10.48550/arXiv.2011.06844>.