National College of Ireland

Programming for AI, MSCAI1, MSCAI1B
Release Date: Monday 25th Nov, 2024
Due Date: Monday 6th Jan, 2025
Shreyas Setlur Arun

_____

Terminal Assessment

Weight: The assignment will be marked out of 100 - weightage 70%

TURNITIN: All report submissions will be electronically screened for evidence of academic misconduct (i.e., plagiarism and collusion)

Duration of Continuous Assessment: 42 days

Attachments: Github link containing all the files

1 Introduction

This project is designed to evaluate the learning objectives of the Programming for AI module as outlined below:

LO1	Analyse, compare, contrast and critically evaluate the characteristics of programming languages andenvironments commonly utilised for AI solutions implementation.
LO2	Critically assess the challenges associated with implementing AI solutions for various problems.
LO3	Critically assess methods and practices for software development in order to design and implementAI solutions requirements.
LO4	Evaluate, design and implement AI solutions by using key algorithms, data structures, and relevantprogramming languages

2 Objectives

The objective of this project is to identify and carry out a series of analyses on a collection of data- sets that are somehow related or complement each other, utilising appropriate programming languages, programming environments and database systems.

This will be a team project. Teams should have 3-4

participants. Your project must incorporate the following

elements:

1. Three or four structured/semi-structured datasets must be used.

2. Datasets must be programmatically stored in appropriate database(s) prior to processing.

3. Programmatic pre-processing, transformation, analysis and visualisation of the data.

4. Programmatically storing the processed output data in appropriate databases.

For example, you could use Python to programmatically retrieve a semi-structured dataset (XML or JSON) or web-scraped or streaming data) and store this data into an appropriate database management system.

You could then read these data to-process and transform it, in the process creating some structured datasets for later usage.

Following that you use Python to conduct further analysis on these data to find interesting patterns by applying knowledge gained in this and other modules (e.g., Foundations of AI), and generate visualizations to better present the results.

Each dataset should contain at least 1,000 records. Some appropriate datasets may be found at:

- https://catalog.data.gov/dataset?res_format=XML
- http://aiweb.cs.washington.edu/research/projects/xmltk/xmldata/
- https://data.gov.ie/dataset?res_format=JSON
- https://catalog.data.gov/dataset?res_format=JSON
- https://data.worldbank.org/

A list of other potential sources will be posted on Moodle.

3   Deliverables

Project Report

The objectives, methodology and results of your analysis should be presented in the form of a project report. This report should discuss the programming and data processing challenges that you encounteredand the means and mechanisms you implemented to overcome these challenges.

The report should be around 3,000 words in length (excluding references), should use appropriate academic style and referencing, and be presented in the IEEE conference format. Templates for Microsoft Word and LATEX can be downloaded from the IEEE[1].

The report should contain the following sections:

- **Abstract**
  This should provide a summary of the project objectives, methods and results. Take a look at abstracts from papers in your literature review to get an idea of what constitutes a good/bad abstract.

- **Introduction**
  Here you should provide a short motivation for the project, describe the relevance of the topic and state the objectives of the project. Note that the **proposed analysis should answer a novel question, which should be clearly stated by means of appropriately formed research**

question(s).

- **Related Work**
  In this section, you should summarise relevant academic work that addressed similar problems or guided your decisions. Note that this should be a **critical evaluation**. It should be more than a mere summary of the works and should discuss their limitations and implications.

---

[1]https://www.ieee.org/conferences_events/conferences/publishing/templates.html

- **Methodology**
  This section should contain :

    - A detailed description of the underlying dataset(s) and your justification for choosing them.

    - Full descriptions and justifications of the data processing activities carried out, such as use of APIs, databases, etc.

    - Complete descriptions and justifications of the implemented data processing algorithms.

    - Justifications for the choice of technologies used, such as programming languages, libraries and databases.

    - Diagrams providing a visual overview of the data gathering, processing and analysis flow.

- **Results and Evaluation**
  Here you should present the results of your work, making appropriate use of figures, tables, etc. You should provide evidence of how the project objectives were met, ensuring that you discuss your research findings, their interpretation(s) and implications.

- **Conclusions and Future Work**
  In this section you should detail what others can/could learn from your work. You should discuss your findings in the context of the research question(s) you elicited earlier. You should present the limitations of your work, i.e. this should be a critical self-evaluation. Lastly, you should suggest potential directions for future work. Typically you would describe what you would do differently or how you would extend your work if you had more time.

- **Bibliography**
  Here you should provide a **complete list** of the academic works cited and online materials used in the project. References should be included as in-text citations **using the IEEE citation style**.

Project Presentation

As a team you should create a video presentation (maximum 10 minutes long) that will act as a discussion point for your work. It should be used to provide a discussion on what you did, how you did it, why you did it and what you discovered.

Note that although individual members will be presenting different parts of the video, each member of the team is expected to be able to present all aspects of the work individually and without assistance from other group members, if required.

Code Artefact

You should create a github repository consisting of all the program code, data and system configuration details and the project journal.

Project Journal

Each member of each project team should maintain their own journal over the course of the project. The journal should provide a brief description of each task carried out by the team member, the time spent carrying out the task, and a description of any challenges or difficulties encountered and how they were addressed.

Note: The submission of the journal is mandatory. Zero marks will be awarded to any team member for the entire project if a journal is not submitted.

The individual project journal will be used to weight the marks awarded to each member of the team, as follows:

Weight	Criteria
70%	A poor journal that fails to provide sufficient information on the team member's contribution to the project, the time spent working on the project or the challenges encountered.
80%	An adequate journal that provides rudimentary information on the team member's contribution to the project, the time spent working on the project or the challenges encountered.
90%	A good journal that provides reasonably in-depth information on the team member's contribution to the project, the time spent working on the project or the challenges encountered.
100%	An excellent journal that provides comprehensive information on the team member's contribution to the project, the time spent working on the project or the challenges encountered.

4 Submission

The project carries 70% of the total marks for the module, with a mark of 40% or greater being required to pass.

For the project report, code artefact and project presentation, there should be only **one submission per team**:

- The project report must include the full name of each team member (as per official documents) and their student number. These must be clearly visible on the front page of the report.

- The code artefact should be uploaded to github and shared in the report

- The project presentation must include the full name of each team member (as per the official NCI documents) and their student number. These must be clearly visible at the start of the video. This should be uploaded as a mp4 video to the github repository or shared as a onedrive link

For the project journal, there will be **one submission for each team member**. Again, this should be in PDF format and should include the full name of student (as per NCI official documents) as well as their student number.

Late submissions will not be accepted unless an extension has been requested through NCI 360 and officially approved.

5 Marking

The project will be marked according to the grading rubric provided at the end of this document.

6 Academic Integrity

Any written work created by others must be properly cited and should be paraphrased or summarised where possible, otherwise it should be included in quotes. Figures not created by you should include an acknowledgement detailing the name(s) of the creator(s). Code found on the internet should not be claimed as your own, but instead a comment should be included in the source code indicating where you obtained it.

Students are strongly advised to familiarise themselves with the Guide to Academic Integrity produced by the NCI Library[2].

> **Note:** All submissions will be electronically screened for evidence of academic misconduct, e.g. plagiarism, collusion and misrepresentation. Any submission showing evidence of such misconduct will be referred to the college's academic misconduct committee for disciplinary action.

---

[2]https://libguides.ncirl.ie/academicintegrity

Grading Rubric – Programming for AI Project

Criterion	Solid H1	H1	H2.1	H2.2	Pass	Fail
Project Objectives **(10%)**	Very challenging project objectives are exceptionally well presented, fully met and thoroughly discussed	Challenging project objectives are well presented, are fully met and thoroughly discussed.	Reasonable project objectives are well presented, fully met and ad- equately discussed.	Reasonable project objectives are clear, are mostly met and adequately discussed.	The objectives are clear, if unambitious and are at least partially met and briefly discussed.	The objectives of the project are unclear, have not been discussed. It is not possible to discern if the objectives have been met.
Literature Review **(10%)**	An excellent critical analysis of substantive and highly relevant literature.	A very good critical analysis of substantive and relevant literature.	A good analysis of relevant literature. The critical analysis aspect could be somewhat stronger.	An adequate analysis of mostly relevant literat- ure. The critical ana- lysis aspect could be significantly stronger.	A limited analysis of some relevant literature but it lacks evidence of understanding.	Little or no relevant literature reviewed. Very limited evidence of understanding.
Data Complexity and Handling **(20%)**	The datasets have been well prepared and meaningfully explored. All datasets were stored in appropriate databases before and after processing. At least two datasets have a high degree of com- plexity. All parts of the analysis are automated within a single process control flow.	The datasets have been well prepared and meaningfully explored. All datasets were stored in appropriate databases before and after processing. At least two datasets have a high degree of complexity.	The datasets have been well prepared and ex- plored. At least one dataset was stored in an appropriate data- base. At least one data- set has a high degree of complexity.	The datasets have been appropriately prepared for analysis. At least one dataset was stored in an appropriate data- base. At least one of the datasets is non- trivial.	The datasets were appropriately handled given the objectives. The use of databases is very basic and some inappropriate choices may be evident. The datasets are somewhat trivial.	Only one somewhat trivial dataset was used. No database was used to store the datasets. No obvious development was carried out.
Data Processing Implementation **(30%)**	The data processing algorithms used play a well conceived and essential role in meeting the project objectives. The implementation significantly exceeds the stated minimum requirements.	The data processing algorithms used play a well conceived and essential role in meeting the project objectives. Multiple data processing techniques / languages were employed.	The use of data processing algorithms is well-thought and appropriate for the project objectives. Comprehensive use of at least one data programming language and multiple techniques.	The use of data processing algorithms is meaningful and appropriate for the project objectives. There is evidence of appropriate use of at least one data programming language and a small number of appropriate techniques.	Appropriate but basic use of data processing algorithms. Basic use of data programming languages and a limited number of techniques.	Poor or no implementation. If an implementation is provided, it demonstrates inappropriate use of data processing algorithms.

Grading Rubric (continued)

Criterion	Solid H1	H1	H2.1	H2.2	Pass	Fail
Results and Conclusions **(10%)**	Three or more insightful findings are excellently presented and thoroughly discussed in the context of the domain using appropriate references to prior work.	Three or more interesting and non-arbitrary findings are presented and thoroughly discussed the context of the domain using appropriate references to prior work.	Three or more interesting non-arbitrary findings are presented and thoroughly discussed.	Two or more interesting non-arbitrary findings are presented and appropriately discussed.	Two or more interesting non-arbitrary findings are presented but are poorly discussed.	Little to no non-arbitrary results and/or findings are presented.
Quality of Writing **(20%)**	Exceptionally well written, with no language errors. All figures are well conceived, readable and correctly captioned. The IEEE template is strictly adhered to. The report does not exceed the length limits. All references are appropriately and correctly used.	Well written, with no significant language errors. All figures are well conceived, readable and appropriately captioned. The IEEE template is adhered to. The report does not exceed the length limits. References are appropriately and correctly used.	Well written, but has a few significant language or style errors. Figures are well presen- ted. The IEEE tem- plate and length limit are adhered to. Refer- ences are complete and correctly used.	Adequately written. but as a few significant language and/or style errors. Some figures are may be hard to read. The IEEE tem- plate and length limit are mostly adhered to. References are com- plete, and correctly used.	Adequately written, with some significant language and/or style errors. Figures may be hard to read or presented in a subop- timal manner. The IEEE template may not have been followed. References are mostly complete and correctly used.	Poorly written and littered with typo- graphical errors and/or poor use of English. The IEEE template was not used. Figures may be hard to read. References (if any) are largely incomplete.