U.PORTO FC FACULDADE DE CIÊNCIAS UNIVERSIDADE DO PORTO / U.PORTO FEUP FACULDADE DE ENGENHARIA UNIVERSIDADE DO PORTO	**Master Degree in Artificial Intelligence** **Artificial Intelligence and Society**	**2025/2026** **1st Year** **1st Semester**
INSTRUCTOR: Miriam Seoane Santos		

Individual Assignment Guidelines

Submission Deadline: Jan 5

Objective

The goal of this individual assignment is to apply and integrate multiple concepts on developing Responsible Artificial Intelligence systems in a small but focused technical study. You will conduct an empirical study that explores the **intersection between at least two topics covered in this course:** *data-centric AI, data complexity, imbalanced data, bias and fairness, missing data, data explainability, data privacy, and synthetic data*.

This assignment encourages you to:

- Extend and demonstrate your expertise on the topics discussed in the course.
- Experiment with the technical methods introduced in class.
- Reflect on the social and ethical implications of your findings.
- Communicate your results clearly and effectively in a structured report.

The deliverables of this assignment are a **written report up to 6 pages and your reproducible experiments**.

Instructions

- You must work **individually**. However, you are encouraged to discuss your topic and approaches with your colleagues and the instructor to receive feedback or work through any bottlenecks.
- Your project should combine **at least two course topics**.
- You may use **real-world publicly available datasets** (e.g., UCI, Kaggle, OpenML) or **synthetic data** that you create to illustrate a concept.
- You can **reuse code** from class tutorials or other authors (e.g., GitHub repositories, papers, etc.), but your work should **introduce some novelty**, adaptation, or meaningful extension to previous work. Simply replicating previous works with minor twists (e.g., changing datasets) is not sufficient.
- Please properly **cite all external sources** you use: code, datasets, tools, or any other(s).

More details are given below regarding **Evaluation Criteria**, **Deliverables**, and **Project Ideas**.

Evaluation Criteria

Your assignment will be evaluated based on the following criteria:

- **Creativity and Problem Definition (3 points)**: Ability to integrate the course topics and define a clear and well-motivated question for investigation; originality in framing the approach to the analysis; creativity in the used tools.
- **Quality of Experiments (7 points)**: Quality and thoroughness of the experiments designed to address the problem; clear motivation and explanation of used approaches, proper implementation of methods.
- **Reflections and Insights (7 points)**: Depth of analysis and critical reflection on the obtained results; ability to interpret its implications, and discusion of limitations and future challenges.
- **Clarity and Organization of Deliverables (2 points)**: Writing and logical flow of the report and structure of the code/data and experiments.
- **Bibliography (1 point)**: Quality of the sources and tools and their relevance to the analysis.

Deliverables

The deliverables of this assignment are a **written report**, along with your **reproducible experiments (data and code)**.

Report Structure

Your report should have **up to 6 pages** using the template presented here. You are free to modify the structure of your report as it fits your needs. However, a clear report generally includes the following:

- **Title and Abstract:** a brief summary of your project and key findings.
- **Introduction and Motivation:** clearly identify the problem you are addressing and why it is relevant from a socio-technical perspective.
- **Background and Related Work:** provide an overview of the methods, techniques or tools used in your work and describe how others have tried to contribute to the problem you are studying.
- **Methodology or Experimental Setup:** describe your experimental framework or pipeline, present your dataset(s), preprocessing techniques (if any), and overall technical approach. You can include pseudo-code, schemas, or other visuals that you find relevant.
- **Results and Discussion:** Present your main findings, supported by figures or tables. Discuss limitations and ethical reflections.
- **Conclusions and Future Work:** Summarize the key takeaways from your work and directions for future exploration.
- **References:** Include references to papers, code, or any other resources that support your work.

Optionally, **you can include an Appendix** with tables, images or other elements that complement your work.

Data and Code

You should structure your pipeline in a way that it can be reproduced by others. You can deliver your data and code in the following ways:

- **.zip file on Moodle:** During the project submission, include a .zip file with your data, a Jupyter Notebook (or .py project) containing the complete code, and a .txt file with instructions on how to reproduce your results.

OR

- **GitHub repository:** Alternatively, you can push your code, data, and documentation to a public GitHub repository. In such case, please include a url in your report. Also, include a short README on your repository explaining how to reproduce your experiments.

Project Ideas

The following projects are simply some ideas that you can build upon. Each integrates two or more topics from the course, and they serve as an example of what type of analysis you can explore in your work:

A. Data profiling is a useful tool to get an overall view of our data and detect possible sensitive information it might contain. Does removing sensitive features alleviate discrimination? How about proxies? Explore some possible techniques for proxy identification.

B. Can data complexity hint at potential unfair situations? Some recent works conceptualize ways in which meta-information can help identify bias in sensitive groups. Produce a proof-of-concept with some experiments of your own. Is this a reasonable claim?

C. The field of imbalanced learning is tighly linked to the field of fairness. Can standard rebalancing techniques improve fairness metrics across sensitive groups? What are their current limitations? How can they be improved? Propose your own fairness-aware resampling method.

D. How can explainability techniques be used to indentify which features most influence biased predictions? How can they help mitigate or alleviate the biases?

E. When data is unavailable or missing, a common strategy is performing data imputation. Explore how different techniques affect fairness or explanations for predictions across different demographic groups.

F. Combine profiling and missing data analysis to characterize systematic missingness. Are missing values random or related to specific groups or attributes? Propose a method to visualize and report missingness as part of a transparent data audit.

G. Can a model be both private and fair? Test whether applying privacy-preserving transformations impacts group fairness.

H. One key challenge we currently face is how to operationalise Responsible AI. Try to design an "ethical auditing tool" workflow that includes profiling, fairness evaluation, and explainability reporting.

I. Generate synthetic datasets to replace or augment real data. Measure how well synthetic data preserves key properties of the original (e.g., statistical diversity, feature correlations, or predictive performance). What are the trade-offs between fidelity and performance?

J. Create controlled synthetic datasets to systematically test how models respond to known data issues (e.g., bias, noise, imbalance, missing data, etc.). Use the results to evaluate model robustness, explainability, or other aspects. Can synthetic data serve as a "sandbox" for auditing AI behavior before deploying on real data?

This assignment is an opportunity for you to connect technical practice with critical thinking, and to explore how seemingly small preprocessing or modeling decisions can have broader ethical and societal consequences. Aim for creativity and rigor! Work on a question that genuinely interests you and use the tools and concepts from this course to reason deeply about it.