**Group Project**

**Project:** **Develop an AI Solution with Responsible AI Principles**

**Submission Date:** December 2, 2024.

In this group project, students will design, develop, and evaluate a simple AI model or system that addresses a real-world problem while incorporating responsible AI principles. The project’s objective is to create an AI solution that addresses fairness, transparency, accountability, and privacy.

In the **Problem Definition** phase, students will choose a real-world problem that AI could address, such as a predictive healthcare model, a recommendation system, or an AI for social good (e.g., disaster relief). They should define the specific objectives of their AI solution, outline how it benefits the target audience, and identify potential ethical issues associated with this type of AI system, such as bias, lack of transparency, or privacy concerns.

In the **Design Phase: Integrating Responsible AI Principles**, students will focus on four main responsible AI principles. For **fairness**, they will define how to ensure that the model avoids bias in training data and fairly represents all relevant groups. For **transparency**, they will outline an approach to make model outputs interpretable and clear in non-technical language. **Accountability** will involve defining how accountability will be maintained for the model’s outputs, including a system for logging decisions. Finally, in addressing **privacy**, students will identify the data they will use, ensuring that privacy concerns are minimized through practices such as anonymizing data and preventing unauthorized access.

Next, in **Model Development and Initial Testing**, students will develop a basic AI model, ensuring that the design choices reflect responsible AI principles. They will train the model on a small, curated dataset, conduct initial tests to evaluate model performance, and check for issues related to fairness, bias, and transparency. All development stages, including data sources, preprocessing methods, and ethical concerns, should be documented.

In the **Evaluation and Improvement** phase, each team will evaluate the model’s performance using standard metrics (e.g., accuracy) and responsible AI metrics (e.g., fairness, interpretability). They will identify any persistent issues with fairness, transparency, or privacy risks, propose at least two improvements to address these issues, and implement one if time permits.

For the final phase, each team will give a **Presentation** (10 minutes). The presentation should summarize the project, highlighting the incorporation of responsible AI principles and challenges faced.

Teams will be assessed on problem definition and ethical considerations (20%), integration of responsible AI principles (30%), technical implementation quality (20%), evaluation and improvement suggestions (20%), and report and presentation quality (10%)