Case Study Activity: Unveiling Bias in Machine Learning

Objective: The objective of this activity is to highlight the ethical implications of biased performance in machine learning models due to unrepresentative or non-diverse data. Students should be able to identify the potential sources of bias and propose methods to mitigate it.

Preparation:

1. Divide the students into groups of 4-5.
2. Provide each group with a case study brief, which includes a description of a machine learning model, the data used to train it, and the context in which it is being used.

Case Study Brief:

Imagine you are a team of data scientists working for HealthAI, a startup that aims to predict the likelihood of diabetes based on a person's lifestyle and genetic factors. The model is trained using a dataset collected from an affluent, predominantly Caucasian neighborhood in a city. The company plans to roll out the product nationally.

Tasks:

1. Identify Potential Bias: Discuss in your group and identify potential sources of bias in this scenario. Consider the demographic, socio-economic, and other factors related to the data used for training the model.
2. Impact Analysis: Analyze and articulate how this bias might affect the performance of the model when it is used nationally. Who could be disadvantaged by this bias? How could it potentially harm individuals or groups?
3. Mitigation Strategies: Develop strategies to mitigate the identified bias. Consider methods for improving data collection, pre-processing techniques, choice of model, post-hoc analysis, etc.
4. Ethical Considerations: Discuss the ethical responsibility of HealthAI in this context. What steps should the company take to ensure its model does not unfairly disadvantage or harm certain groups?

Deliverables:

Each group should prepare a brief report or presentation summarizing their findings and proposals. This should include:

1. A summary of the identified biases and potential impacts.
2. Detailed mitigation strategies and their expected improvements.
3. Ethical considerations and recommendations for HealthAI.

Discussion:

After the presentations, facilitate a discussion about the broader ethical implications of bias in machine learning models. Discuss potential systemic biases that could be introduced in the model development process and the responsibilities of data scientists and organizations in preventing and mitigating bias.