Ethical Reflection - Task 3

Ethical Reflection

1. Potential Biases in the Dataset

The breast cancer diagnosis dataset used in Task 3, such as the one provided in the IUSS 23-24 Kaggle competition, may contain several hidden biases that can affect the model's fairness and reliability when used in real-world clinical settings.

Underrepresented Groups:

The dataset may have a majority of patient data from specific demographics (e.g., White women), while minority populations (e.g., Black or Asian women) are underrepresented. This can lead to a model that performs worse on those minority groups, leading to unequal diagnostic accuracy.

Imaging Source Bias:

If the dataset is collected from a single hospital or imaging system, it might not generalize well across different hospitals or equipment types.

Severity Imbalance:

There might be an uneven distribution of early-stage and late-stage breast cancer cases. The model might be good at detecting early-stage but poor on advanced cases. 2. How Fairness Tools like IBM AI Fairness 360 Can Help

IBM AI Fairness 360 (AIF360) is a toolkit designed to detect, understand, and mitigate bias in machine learning models.

Bias Detection:

AIF360 can calculate metrics like disparate impact and statistical parity difference to assess whether the model treats all groups fairly.

Bias Mitigation:

Techniques such as reweighing and optimized preprocessing can balance the data before training. In-processing methods (e.g., prejudice remover) or post-processing methods (e.g., equalized odds) can

adjust the model or predictions for fairness.

Dataset Auditing:

AIF360 can help identify which groups are underrepresented and recommend actions to address the imbalance, such as collecting more data or applying sampling strategies.

Transparency:

AIF360 provides fairness reports that can be shared with stakeholders to ensure accountability and ethical deployment of AI in healthcare settings.

Conclusion:

Deploying a predictive model for breast cancer diagnosis in a company without considering dataset biases can lead to unequal healthcare outcomes. Fairness tools like IBM AI Fairness 360 ensure that the model remains inclusive, safe, and effective across all patient groups.