

## Ethical Reflection on Predictive Resource Allocation Model

### 1. Potential Biases in the Dataset

Although the Breast Cancer Wisconsin dataset is widely used for research and benchmarking, deploying a predictive model based on this dataset in a real-world company context (e.g., for resource allocation or issue prioritization) raises key ethical concerns:

#### ○ a. Representation Bias

The dataset was collected from a specific geographic and demographic group (patients from one hospital or region).

This can result in underrepresentation of diverse populations (e.g., racial, ethnic, age, gender variations).

If applied broadly, the model may generalize poorly and misprioritize resource allocation for underrepresented groups.

#### b. Labeling Bias

The priority variable in Task 3 was simulated from the mean radius feature.

While this works for practice, it assumes that tumor size directly correlates with issue priority — which may not reflect real-world clinical or organizational decision-making.

#### c. Automation Bias

Decision-makers might over-rely on the model, ignoring critical context that algorithms can't capture (e.g., a benign case in a high-risk patient might still need fast action).

Without proper oversight, resource allocation based solely on model output could amplify disparities.

### 2. Addressing Bias with AI Fairness Tools

Tools like IBM AI Fairness 360 (AIF360) provide a framework to detect, measure, and mitigate bias in machine learning models.

#### a. Bias Detection

AIF360 can evaluate fairness metrics such as disparate impact, equal opportunity, or statistical parity.

Example: If the model assigns higher priority more often to one group (e.g., younger patients), it will flag this as a disparity.

#### b. Preprocessing Techniques

Tools in AIF360 can rebalance datasets before training using techniques like reweighting or resampling to ensure that all groups are fairly represented.

#### c. In-processing Fairness Algorithms

Modify the learning algorithm to promote fairer decisions — such as the Adversarial Debiasing model or Prejudice Remover Regularizer.

#### d. Post-processing Corrections

After model predictions, AIF360 can apply fairness-enhancing techniques to adjust outcomes — e.g., equalized odds post-processing.

## Conclusion

Ethical AI practices are crucial when predictive models are used for critical decisions like resource allocation. While our Task 3 model is technically accurate, fairness tools like IBM AIF360 are essential to:

Detect hidden bias,

Ensure fair outcomes for all user groups,

Maintain trust and accountability.