

### Problem Statement:

This project addresses the issue of socioeconomic bias in a machine learning model used for loan approval predictions. The model may favor individuals from higher socioeconomic backgrounds (based on income and education level) over those from lower socioeconomic backgrounds, even when both groups have similar financial qualifications, such as credit score and employment status. This bias leads to unequal treatment, where lower-income individuals face more difficulty in obtaining loans despite being equally qualified.

The human value being ignored is fairness in access to financial resources, the principle of Equal Opportunity. By failing to treat all applicants fairly, the model maintains systemic inequalities and disadvantages lower-income individuals. The aim of this project is to ensure that the model's predictions follow Equal Opportunity, providing an equal chance of loan approval to applicants with similar financial qualifications, regardless of their socioeconomic status.

### AI System Overview:

The AI model will be trained on a loan application dataset that includes features such as:

- Credit score
- Annual income
- Employment status
- Educational background
- Loan amount requested
- Loan approval status

Socioeconomic status (SES) is represented through variables such as income and education level, which are critical in identifying bias in loan approvals.

### Computation (Deep Learning Architecture and Training Strategy):

The model used is a feedforward neural network designed for binary classification, predicting whether a loan will be approved or denied. The network architecture will consist of:

- Input layer: Processes the features from the loan application dataset.
- Hidden layers: Apply non-linear transformations to capture complex patterns in the data.
- Output layer: Outputs the binary prediction (approve/deny).

The model will be trained using binary cross-entropy as the loss function and stochastic gradient descent (SGD) as the optimization algorithm. The training strategy includes:

- A train-validation-test split (80% training, 10% validation, 10% test),
- Batch normalization to improve learning stability and reduce overfitting,
- Dropout in the hidden layers to promote generalization.

### Human Value Being Ignored:

The human value being overlooked is fairness, specifically Equal Opportunity, which means that individuals with similar financial profiles (e.g., credit score, employment status) should have an equal chance of loan approval, regardless of their socioeconomic background. Fairness is essential to prevent the systematic disadvantage of individuals from lower socioeconomic backgrounds in accessing financial resources.

If the model disproportionately favors higher SES applicants, it reinforces existing socioeconomic inequalities by denying lower-income individuals access to credit and opportunities for financial mobility.

#### Quantification of the Human Value:

Fairness will be quantified using the true positive rate (TPR), which measures the percentage of qualified applicants who are correctly approved for loans. Equal Opportunity is violated if the true positive rate for lower SES applicants is significantly lower than that for higher SES applicants.

For example, if qualified lower-income applicants are approved at a substantially lower rate than their higher-income counterparts, the disparity will provide a measurable indicator of bias in the model's decision-making. The magnitude of this gap will quantify the extent to which fairness has been compromised.