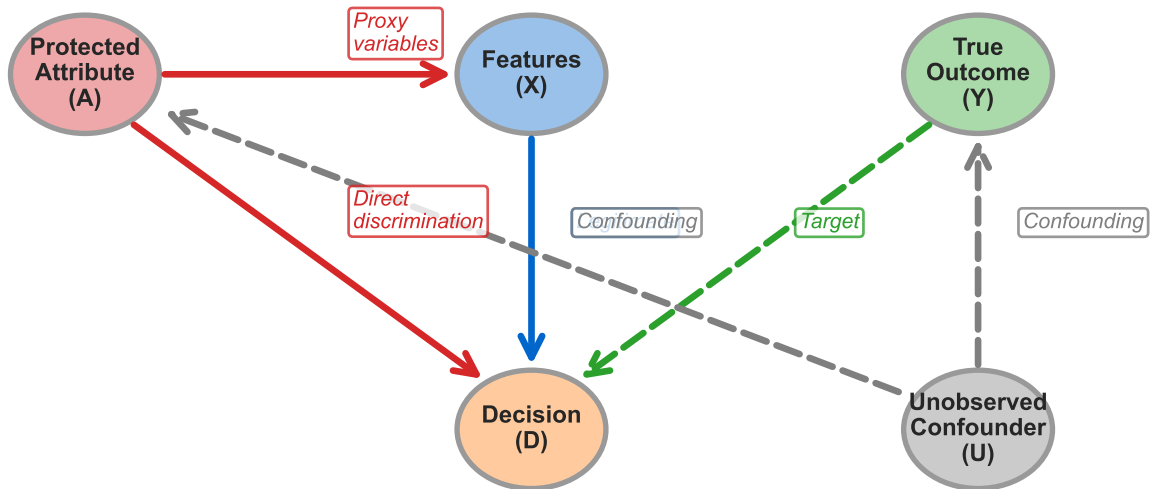


Causal DAG: Direct vs Indirect Discrimination



Direct path: $A \rightarrow D$ (red solid)

Indirect path: $A \rightarrow X \rightarrow D$ (red + blue)

Legitimate path: $X \rightarrow D$ (blue, no A involved)

Counterfactual fairness requires blocking $A \rightarrow D$ and $A \rightarrow X \rightarrow D$ paths