Interpretable ML for Clinical Adoption

⚠️ **Black Box Problem:** Clinicians won't trust models they can't understand

Glass Box Models

• Linear/Logistic Regression
• Decision Trees
• GAMs (Generalized Additive Models)
• Rule-based systems

Inherently interpretable

Post-hoc Explanations

• Feature Importance (RF, XGBoost)
• Partial Dependence Plots
• LIME (Local Interpretable)
• SHAP Values

Explain black boxes

Clinical Acceptance Requires: Model explanation + Clinical validation + Physician trust