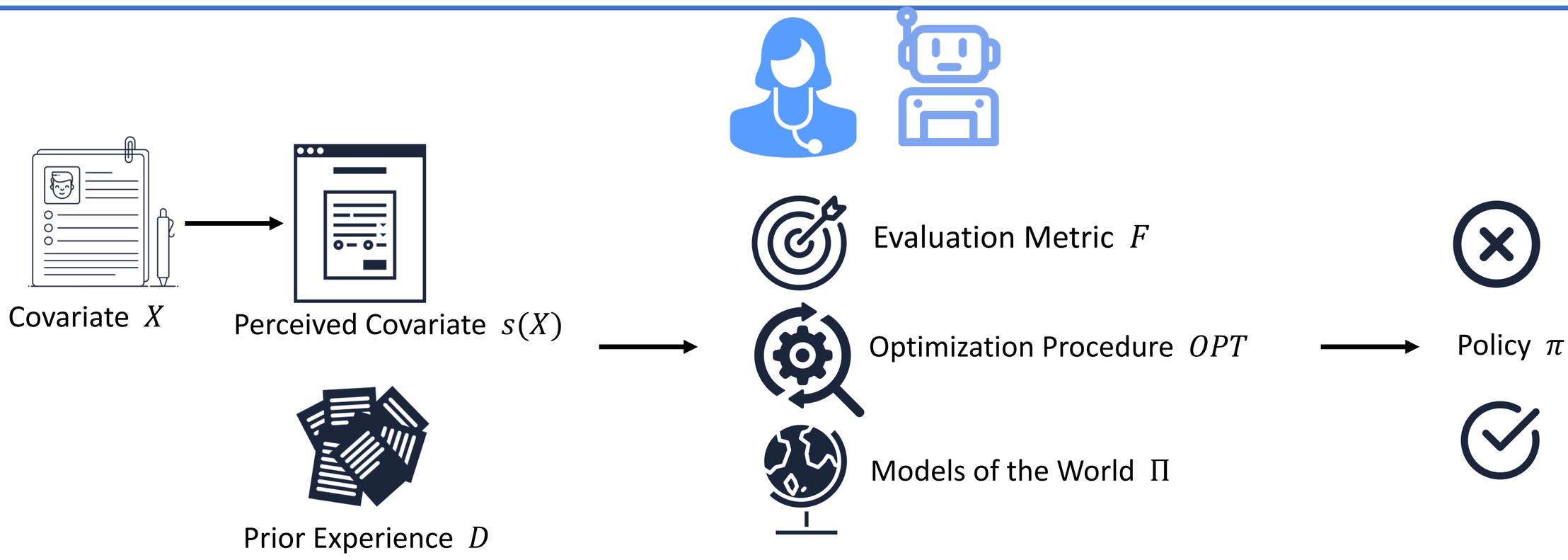
A Unifying Framework for Combining Complementary Strengths of Humans and ML toward Better Predictive Decision-Making

Liu Leqi*, Charvi Rastogi*, Kenneth Holstein, Hoda Heidari
Carnegie Mellon University

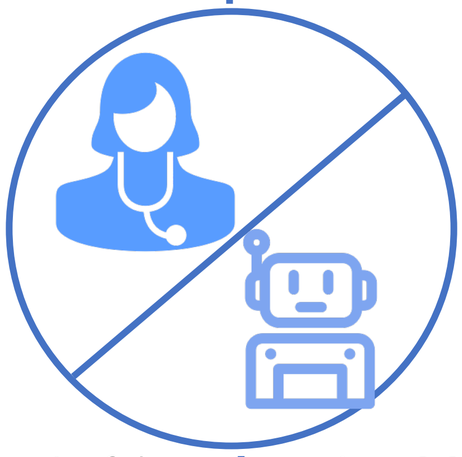


TASK DEFINITION

1. Objective:
   - Humans: comprehensive and holistic objectives in decision-making, such as human relations, fairness, future outcomes.
   - ML: mathematically tractable objective minimizing expected error in supervised learning, or maximizing rewards.
   - Complex v/s simple objectives. Multiple v/s single objectives.
2. Construct of Interest:
   - Social contextual decision-making involves unobservable theoretical constructs such as risk of recidivism, risk of surgery, teacher effectiveness.
   - ML: Bias in measurement of observable properties lead to unfairness and social harms.

INPUT

1. Access to Different Information:
   - Hard to codify all information, doctors observe the physical presentation of patients, their support system, judges observe the predisposition of defendants.
2. Nature of Past Experiences:
   - Human experience and learning amassed over long time; wide and short data.
   - ML has large number of prior instances for specific task with limited features; narrow and long data.
   - Rich contextual experiential learning v/s case-based input-output based learning.



OUTPUT

1. Explaining the Decision
   - Humans: generate coherent explanations that are meaningful to other humans.
   - Humans: explanations are contrastive, selected in a biased manner, social and contextual.
   - ML: Decisions made by models are traceable, but not understood by laypeople.
2. Uncertainty Communication
   - Humans: find it difficult to quantify uncertainty.
   - Humans: different people have different calibration of uncertainty.
   - ML: many uncertainty quantification methods have been studied, research is ongoing.
3. Output Consistency
   - Humans: Judgements by human show random inconsistency independent of the task at hand (time of day, external perturbations). Not true for ML.
4. Time Efficiency
   - Humans are much slower than ML at producing decisions.
   - ML can produce a large volume of decisions together.

INTERNAL PROCESSING

1. Models of the World
   - Humans: Rich mental models encoding complex beliefs about causal mechanisms.
   - ML: Tractable hypothesis class of statistical models.
2. Choosing among Models of the World
   - Human: Picking a model using unknown heuristics, satisficing behavior.
   - ML: First-order optimization requiring extensive computation.
3. Internal Processing and Perception
   - Computational capacity of humans is lower than ML. Cognitive boundedness gives lower grained perception by humans.
   - Different perceptual biases displayed.
   - Humans have a causal perception of data whereas machines have a statistical perception.

How do we certify the existence of complementarity? Find π^*, the optimal aggregated policy wrt F. Complementarity exists if and only if $F(\pi^*) > max\{F(\pi_H), F(\pi_M)\}$.