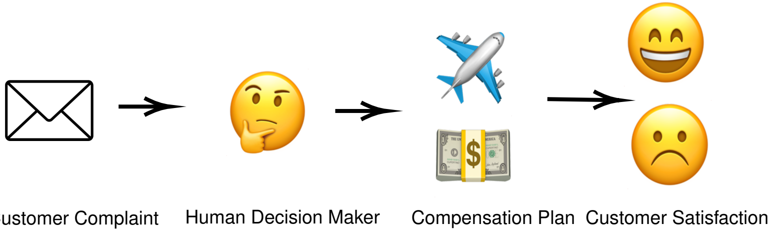
Confounding-Robust Policy Learning with Human-AI Teams

Mingzhang Yin
Department of Marketing
Artificial Intelligence Academic Initiative
University of Florida

Human Systems

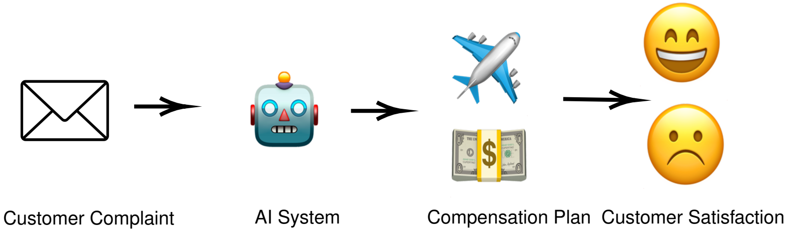
- Decision-making data are often collected from human experts
- Customer support system as an example:



- However, human decisions are time consuming and hard to scale

Automated AI Systems

- Many companies adopt AI systems for automated decision-making



- The AI policy $\pi(T \mid X)$ generates treatment / decision T given input features X
- However, AI systems' performance heavily rely on the model assumptions and the training data

Deferral Collaboration of Human-AI

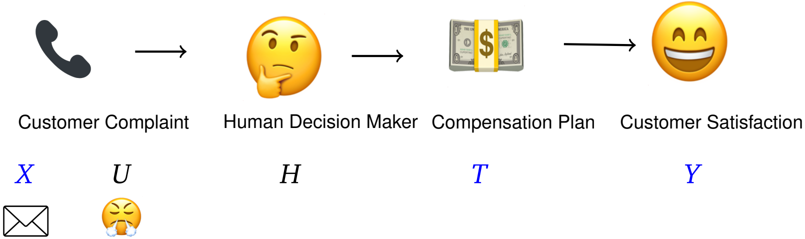
- We build a deferral collaboration system (Madras, Pitassi, and Zemel 2018) to achieve Human-AI complementarity
- Jointly learns:
  - A routing algorithm $\phi(X) : X \to [0, 1]$ — probability of assigning task to a human.
  - An algorithmic policy $\pi(X) : X \to \Delta_m$ — distribution over m possible treatments.
- An ideal Human-AI system outperforms both human-only and AI-only approaches

Challenges of Information Asymmetry

- Human decision-makers usually have access to both recorded information (X) and unrecorded information (U)
- AI algorithms are trained only on recorded information (X)
- Example in customer support:
  - X: travel purpose, delay duration, flight experience
  - U: caller emotion, travel context nuances
- U leads to unmeasured confounding problem, as it influences both the treatment (Compensation Plan) and the outcome (Customer Satisfaction)
- Goal: design a Human-AI system that is robust to unmeasured U

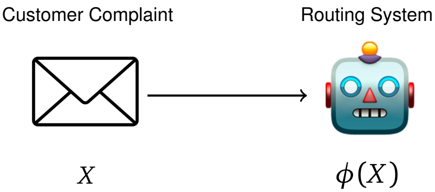
Set-up: Policy Learning with Observational Data

- The data $\{X_i, T_i, Y_i\}_{i=1}^{N}$ are collected from past decisions by human experts

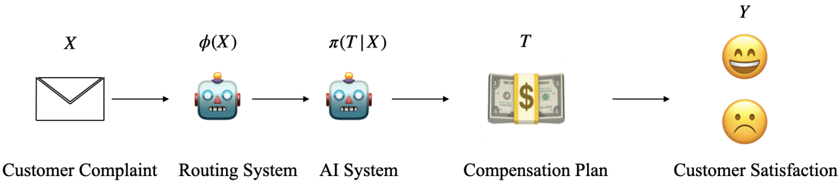


- The treatment is generated by an unknown behavior policy $\pi_0(T \mid X, U) = p(T \mid X, U)$ depending on U

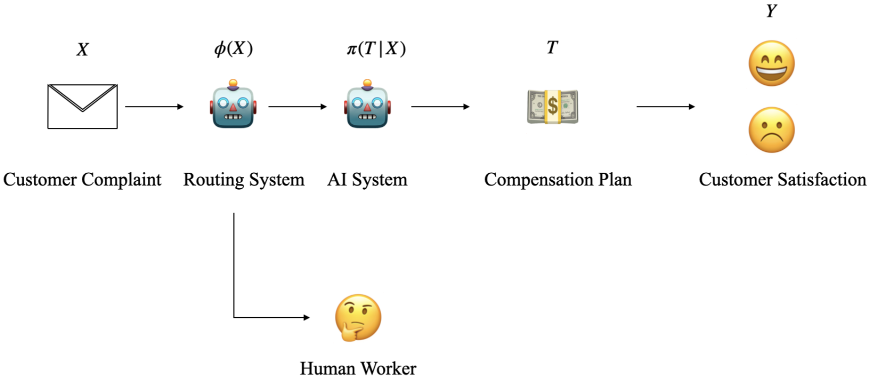
Framework Overview



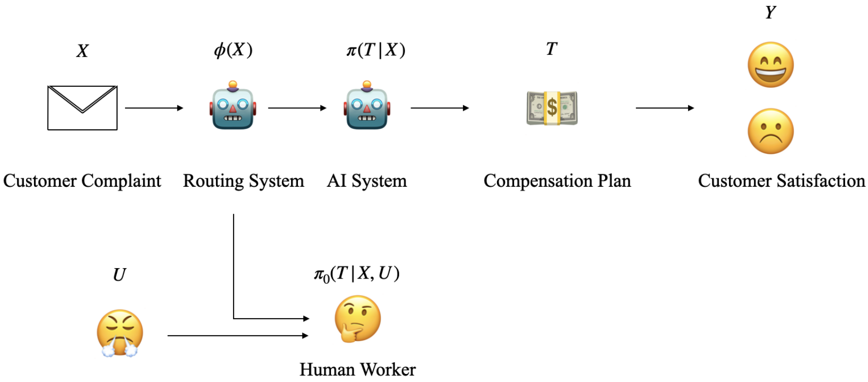
Framework Overview



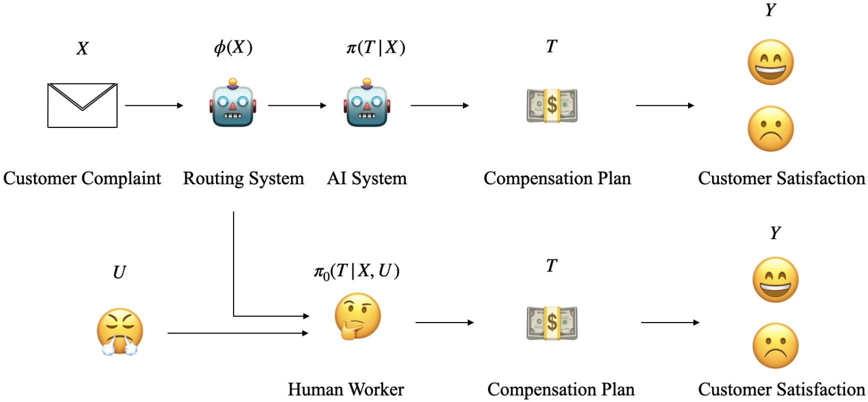
Framework Overview



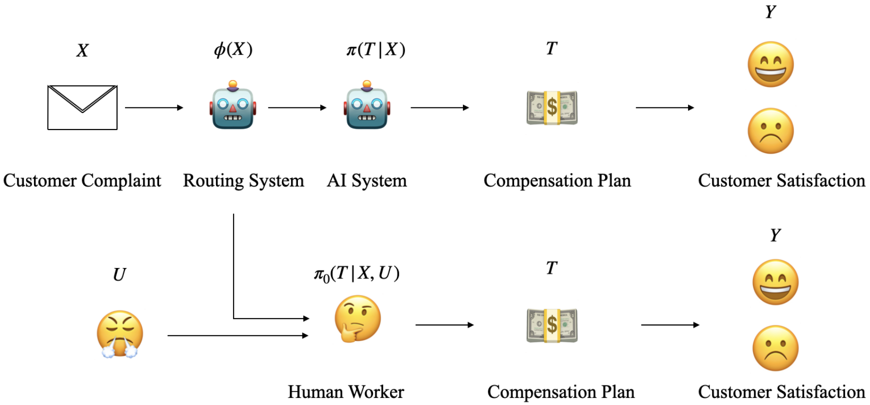
Framework Overview



Framework Overview



Framework Overview



How to optimize both the Routing system ($\phi(X)$) and the AI system ($\pi(T \mid X)$) on data $\{X_i, A_i, Y_i\}_{i=1}^{N}$ sampled from the behavior policy $\pi_0(T \mid X, U)$?

Learning Complementary Policies

If we have access to both recorded X and unobserved U, the system can be optimized by minimizing the policy regret:

$$\min_{\pi,\phi} R(\pi,\phi;\pi_c) = \mathbb{E}\left[\phi(X)\big(Y + C(X)\big)\right] + \sum_{t=0}^{m-1} \frac{\mathbb{E}\left[\frac{\mathbb{I}(T=t)}{\pi_0(T \mid X,U)} Y\left((1-\phi(X))\pi(T \mid X) - \pi_c(t \mid X)\right)\right]}{\mathbb{E}\left[\frac{\mathbb{I}(T=t)}{\pi_0(T \mid X,U)}\right]}$$

↑ Human performance ↑ AI policy performance

- Y: Cost (e.g., dissatisfaction)
- $C(X)$: Cost of using a human decision-maker
- $\phi(X)$: Routing probability to humans
- π: AI policy
- π_c: Baseline policy (e.g., default)
- $\pi_0(T|X,U)$: Behavior policy for data collection

Accounting for Unobserved Confounding

- We adopt the Marginal Sensitivity Model (Tan, 2006) to model confounding:

$$\Gamma^{-1} \leq \frac{(1-\tilde{\pi}_0(T \mid X))\pi_0(T \mid X, U)}{\tilde{\pi}_0(T \mid X)(1-\pi_0(T \mid X, U))} \leq \Gamma$$

- Γ: measures confounding strength — obtained from domain knowledge or estimated from data.
- We optimize the Human-AI system by solving a robust min-max problem under this model:

$$\min_{\phi,\pi} \max_{W} \frac{1}{n} \sum_{i=1}^{n} \phi(X_i)(Y_i + C(X_i)) + \sum_{t=0}^{m-1} \frac{\frac{1}{n}\sum_i I(T_i = t)[(1-\phi(X_i))\pi(T_i \mid X_i) - \pi_c(T_i \mid X_i)]W_i Y_i}{\frac{1}{n}\sum_i I(T_i = t) W_i}$$

$$\text{s.t.} \quad 1 + \Gamma^{-1}(\tilde{W}_i - 1) \leq W_i \leq 1 + \Gamma(\tilde{W}_i - 1),\ \tilde{W}_i = \tilde{\pi}_0(T_i \mid X_i)^{-1}$$

Theoretical Results

What instances should be routed to humans?

- Route to human ($\phi(X) = 1$) when

$$\mathbb{E}_{U\sim P(U|X),T\sim\pi_0(T|X,U)}[Y + C(X)|X] \leq \mathbb{E}_{T\sim\pi(T|X)}[Y|X]$$

Risk of human utilizing unobserved information U

Expected risk of routing the instance to the AI

- $C(X)$: Human cost on instance X.
- $\phi(X)$: Probability of routing to humans.
- $\pi(T \mid X)$: Probability of AI policy taking action T.
- $\pi_0(T \mid X, U)$: Human's behavior policy.

- Even with infinite data and powerful models, incorporating human judgment still adds value to the system.

Theoretical Results: Improvement Guarantees

Theorem (Informal)

Suppose outcomes and human costs are bounded, and the behavior policy assigns non-negligible probability to all actions. Let Π *and* Φ *be the classes of AI policy and routing policy. Then, with high probability:*

$$R(\pi, \phi; \pi_c) \leq \hat{R}_n(\pi, \phi; \pi_c) + \mathcal{O}\left(\frac{1}{\sqrt{n}}\right) \textit{Rademacher complexities terms}$$

- When baseline $\pi_c \in \Pi$, the empirical objective $\hat{R}_n(\pi, \phi; \pi_c)$ is guaranteed to be negative
- For sufficiently large n, the regret is guaranteed to improve over the baseline policy

Synthetic Data:

- **Data-generating process** (Kallus and Zhou 2021):

$$\xi \sim \mathrm{Bern}(0.5), \quad X \sim \mathcal{N}\big((2\xi-1)\mu_x,\ I_5\big),$$
$$U = \mathbb{I}[Y(1) < Y(-1)],$$
$$Y(t) = \beta^\top X + \mathbb{I}[t=1]\,\beta_{\text{treat}}^\top X + 0.5\,\alpha\xi\,\mathbb{I}[t=1] + \eta + w\xi + \varepsilon.$$

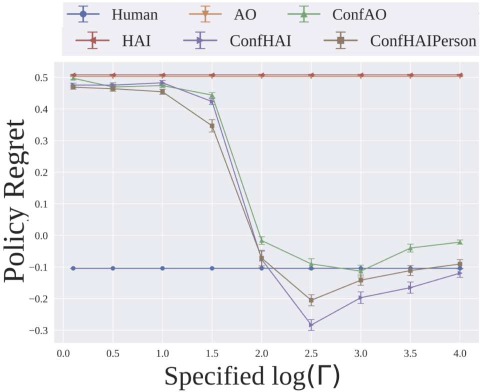
- **Parameters** (fixed):

$$\beta_{\text{treat}}=[1.5, 1, 1.5, 1, 0.5],\ \mu_x=[1, .5, 1, 0, 1],\ \eta=2.5,\ \alpha=-2,\ w=1.5;$$
$$\beta = [0, .75, .5, 0, 1, 0],\ \varepsilon \sim \mathcal{N}(0,1).$$

- **Nominal propensity** $\pi_0(T=1 \mid X) = \sigma(\beta^\top X)$, **True propensity** $\pi_0(T=1 \mid X, U)$

Experiments

- Compare with the state-of-the-art method in human-AI collaboration assuming unconfoundness (HAI) and policy learning with unobserved confounding (ConfAO).
- With correctly or over specified confounding strength, we observe consistent policy improvement.



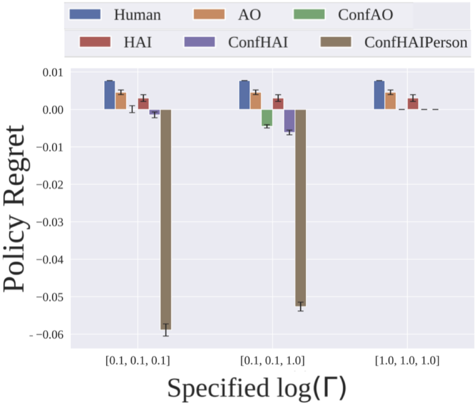
Experiments

Approval of home improvement personal loan (data from LendingClub)

- **Observed features:** Borrower income, credit utilization, debt-to-income ratio, number of delinquencies, employment length, etc.
- **Unrecorded features:** Additional information acquired via in-person applicant interactions
- **Treatment:** Loan approved or denied
- **Outcome:** Whether the borrower defaulted or repaid

Experiments

- ConfHAI improves outcomes over Human-only, Algorithm-only, and standard Human-AI models
- Lower loan default rates through robust human-AI task routing



Conclusion

- We propose a novel algorithm for human-AI collaboration robust to unobserved confounding.
- The system leverages human decision-makers who access additional, unrecorded information.
- We provide theoretical guarantees for policy improvement over both algorithm-only and human-only decision-making.
- More experiments, including healthcare applications, are available in the full paper

See: `https://ojs.aaai.org/index.php/AAAI/article/view/33559` (AAAI 2025),
Joint work with Ruijiang Gao (UT Dallas)

Thank You!