Responsible Uplift Modeling

AAAI Track

Lihi Idan
Texas A&M University
College Station, United States
idanlihii@gmail.com

Ming Li
Nanjing University
Nanjing, China
ming.li8723@gmail.com

ABSTRACT

Automated intervention policies have become highly prevalent within firms, with "algorithmic personalization" techniques at their foundation. These methods leverage individual-level data to decide which groups should be targeted by the firm's policies. While such policies are naturally guided by the multi-dimensional heterogeneity that exists among individuals, relying on some dimensions of such heterogeneity may unintentionally result in biased outcomes for socially-disadvantaged groups.

This work focuses on a particular form of personalization: Uplift Modeling. While research on fairness in algorithmic personalization has been growing in recent years, the broader societal impact of Uplift Modeling has largely been overlooked in previous technical work. We introduce the first in-processing, learning-based method for Fair Uplift Modeling, applicable in both static and dynamic environments. Our Uplift Models are evaluated on real-world datasets, demonstrating promising results.

KEYWORDS

Uplift Modeling; Fairness; Bias

ACM Reference Format:
Lihi Idan and Ming Li. 2025. Responsible Uplift Modeling: AAAI Track. In *Proc. of the 24th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2025), Detroit, Michigan, USA, May 19 – 23, 2025*, IFAAMAS, 8 pages.

1 INTRODUCTION AND KEY IDEAS

Automated targeting policies have become highly prevalent within firms. At the core of such policies are "algorithmic personalization" techniques: a set of automated methodologies that utilize customers' personal data to decide which customer should be assigned a given treatment. While such personalized policies are guided by the multi-dimensional heterogeneity that exists among individuals, relying on some dimensions of such heterogeneity when creating the policy can become a double-edged sword as it might lead to discriminatory behaviors, resulting from the policy having a substantially different impact on different groups of the population. While impact disparities among different groups of the population are considered legitimate if the groups mostly differ in their unprotected attributes such as interests or talents , other attributes such as race and gender are considered protected attributes; a policy disproportionately affecting different groups based on such protected attributes is considered an unethical policy both socially and legally [36].

This work focuses on a specific type of personalization endeavor used extensively in the marketing domain, Uplift Modeling. Intuitively, Uplift Modeling aims at predicting not only an individual's tendency to perform a behavior of interest given her unique profile but rather the *causal effect of a treatment* on the individual's tendency to perform the behavior of interest. Uplift Modeling thus yields a *personalized intervention policy* in which intervention decisions are determined based on the estimated *incremental* impact of the treatment instead of its absolute impact. Our goal is to design Uplift models resulting in intervention policies that are both highly effective, yielding high profits to firms who use them, and at the same minimize outcome disparities resulting from the policies' deployment in heterogeneous environments.

1.1 Disparate Treatment and Disparate Impact

U.S. anti-discrimination law identifies two types of discriminatory behavior. The first, disparate treatment law, seeks to prevent decisions that intentionally rely on protected characteristics [7]. The second, disparate impact law, makes illegal policies that result in unnecessary and unjustified disproportionate effects. [Griggs v. Duke Power Co., 401 U.S. 424 (1971)]. The main difference between the two forms of discriminatory conduct is the stage they are located within the policy's life cycle: while treatment-based disparities are concerned with policymakers' *intentions*, impact-based disparities are concerned with the *effects* of a policy, even if the discriminatory effects are unintentional.

To illustrate the two notions of discrimination, let us consider the following scenario: Amazon is launching a new Prime service. As Amazon often partners with its subsidiaries for marketing its own services, Amazon decides to promote its new service via the following coupon campaign: each individual targeted with a coupon and registers for the new service receives a 20-dollar coupon for purchases made at Amazon's subsidiary, Whole Foods Market. Since African-American individuals tend to buy less online [34], Amazon decides to target only non-African-American individuals with coupons. Such an intervention policy will result in disparate treatment on the race attribute and hence trivially to disparate impact.

Let us consider a different scenario, in which the U.S. extends the Fair Housing Act [13] so it also applies to online retail. In such a case, Amazon can no longer explicitly determine its targeting policy based on customers' ethnicity. Instead, Amazon decides to target only individuals from high-income households, knowing that such individuals are more likely to buy from Amazon compared to those from low-income households [35].

Proc. of the 24th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2025), Y. Vorobeychik, S. Das, A. Nowé (eds.), May 19 – 23, 2025, Detroit, Michigan, USA. © 2025 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org).

Observation: The ethnic composition of high-income households in the U.S. is disproportionately biased towards non-African-American individuals [14].

In such a case, even though the policy does not explicitly undertarget African Americans, the proportion of non-African-Americans using the new service will likely be significantly higher than the proportion of African Americans compared to each group's representation in the overall population. The strong correlation between income level and ethnicity leads to fewer African Americans being targeted with coupons, which most likely will lead to fewer registrations to the new service, resulting in disparate impact.

The former is the most well-researched disparate-impact source, which we refer to as Disproportionate Treatment Allocation: The disproportionate effects are caused by a disproportionate treatment allocation based on an attribute that is strongly correlated with a protected attribute. However, there exists another, more elusive source of disparate impact that is hardly addressed in prior technical work.

Let us consider a third scenario, in which one of Amazon's stakeholders is particularly interested in promoting Amazon's compliance with diversity and inclusion values [40]. Thus, the stakeholder strictly opposes policies resulting in racial disparities. Amazon thus decides to equally target African American individuals and non-African-American individuals (with respect to each group's proportion in the population) hoping that by using such a balanced targeting policy the final outcome, registration to the new service, will be balanced ethnicity-wise.

Observation: Premium grocery stores are missing from African-American neighborhoods [9].

Observation: Proximity has a strong effect on consumers' decision on where to shop [1].

In such a case, though the intervention is proportionally balanced, the proportion of non-African-Americans using the new service will likely be significantly higher than the proportion of African Americans compared to each group's representation in the overall population. The coupon is less attractive to African-American individuals than to those from other ethnic groups, leading to their disproportionate *decision* not to join the new service. The campaign thus results in disparate impact, even though it is disparate-treatment-free. Importantly, the disproportionate effects are caused by what we refer to as *decisional disparities*: systematic differences in *self-selected outcome decisions* made by each demographic group within the treated population.

Using the above illustration, one can identify three different sources of disparate impact in personalized intervention policies: Disproportionate Treatment Allocation, Decisional Disparities within the treated population, and Decisional Disparities within the untreated population (an illustration of the latter was not given due to lack of space). We claim that uplift models aimed at minimizing disparate impact must account for *all* sources of disparate impact including decisional disparities even though those result from customers' self-selected choices and are thus perceived as lying beyond the policymaker's control; this is essential in order to prevent firms from misusing implicit correlations between the nature of the intervention and subjects' preferences in order to systematically discriminate certain demographic groups while still maintaining a neutral-looking policy.

To determine whether a policy results in disparate impact, we use a formal test of disparate impact which typically consists of three key elements.

1. Adverse impact: The policy disproportionately impacts the minority group [New York City Env't Just. All. v. Giuliani, 214 F.3d 65 (2d Cir. 2000)].

2. No Justification: The adverse impact does not have a substantial justification rooted in a legitimate policy goal [Texas Dep't of Hous. & Cmty. Affs. v. Inclusive Communities Project, Inc., 576 U.S. 519 (2015)].

3. Less discriminatory alternative: Even if the disparate impact is justified, the one claiming against the policy will prevail if there is an *alternative feasible policy with less adverse impact on the minority group* [Elston v. Talladega Cnty. Bd. of Educ., 997 F.2d 1394 (11th Cir. 1993)].

This work focuses on the third element: We develop Uplift models for the automatic design of disparate-impact minimizing, personalized intervention policies. Furthermore, we show that such policies not only exist, but are also feasible in the sense of achieving an accuracy level that is comparable to that of more biased, discriminatory policies. By doing that, we present *an alternative feasible policy with less adverse impact on the minority group* as article 3 requires. The uplift models presented in this work are able to account for *all* sources of bias using the same underlying model by departing from the traditional preprocessing and postprocessing methods and instead directly regularizing the predicted outcome during training. By regularizing only the model's predicted outcomes without enforcing *any* constraints on the model's treatment allocation, we are able to account for *all* three disparate-impact sources, both those that stem from disproportionate treatment allocation and those that stem from decisional disparities.

The metric that we use to quantify the level of disparate impact produced by a policy in this work is the "Equalized Relative Factual Outcome" metric, which can be seen as the equivalent of demographic parity used in the disparate-treatment context [15]: An ideal policy, fairness-wise, is a policy that results in each group having a positive factual outcome rate that equals to the group's proportion in the general population. Formal definitions of the above are given in Section 2.

1.2 Our contributions

In this work, we present the first in-processing uplift-modeling approach for minimizing *all sources* of disparate impact, including those that stem from decisional disparities, in both static and dynamic environments. Our unique contributions are as follows:

1. Unlike prior work on algorithmic bias which focuses on traditional predictive modeling, we focus on bias minimization in uplift modeling techniques.

2. Unlike prior work on the use of machine learning for designing fair personalized policies which have traditionally focused on *disparate-treatment* minimization, our work aims at designing new uplift models for minimizing *disparate impact* in personalized intervention policies.

3. Unlike prior works on fair personalized policies that are limited to static policies, we present disparate-impact-minimizing models that support the creation of *dynamic* policies; that is, our models

aim at minimizing the *cumulative* disparate impact in multi-step intervention policies.

2 PROBLEM DESCRIPTION

We consider a setting in which a firm initiates a business process aimed at obtaining a given outcome — known to be positively correlated with a given treatment ("intervention") that can be applied to customers by the firm. However, due to cost-related constraints the firm can only treat a limited number of customers. The firm's goal is thus to create a *personalized intervention policy*: a mapping from customers to treatment assignment, guiding the firm which customers to treat. In order to ensure cost-effective interventions, the firm uses an Uplift Modeling technique for creating the intervention policy, as we describe below.

We consider a set of n customers, each associated with a vector of features $X_i \in \mathcal{X}$. We assume a subset of the features, X_i^s to be protected (*i.e.* it includes attributes such as age and gender). In the next sections, for simplicity of explanation we assume that $|X_i^s| = 1$ and refer to the protected attribute as S, a discrete attribute; our methods can be trivially extended to consider X_i^s of any size and of any type.

Each customer can be assigned one of M treatments, $W_i \in \mathcal{W}$, $|\mathcal{W}| = M$. For simplicity, we consider the traditional case where $M = 2$; our models can be easily extended to the case of an arbitrary M. Each of the n customers is further associated with multiple *potential outcomes* Y_i^j denoting the outcome of customer i when receiving a treatment j. In our simplified setting assuming $M = 2$, we consider only Y_i^0 and Y_i^1: Y_i^0 denotes customer i's outcome when she does not receive the treatment and Y_i^1 denotes customer i's outcome when she receives the treatment.

Our goal is to create an intervention policy $\Pi : \mathcal{X} \rightarrow \mathcal{W}$ which maximizes the firm's utility given a fixed cost, C, denoting the number of customers that can be treated:

$$U = \frac{1}{n} \sum_{i=1}^{n} E[Y_i^{\Pi(X_i)}] \tag{1}$$

A well-known approach for obtaining such policy is building an Uplift Model which estimates τ_i, the conditional average treatment effect (CATE):

$$\tau_i(x) = E[Y_i^1 - Y_i^0 | X_i = x] \tag{2}$$

And using $\hat{\tau}$ as well as C for selecting whom to treat. Assuming only two types of treatments, the common way to estimate τ_i using W_i is:

$$\hat{\tau_i}(x) = E[Y_i | X_i = x, W_i = 1] - E[Y_i | X_i = x, W_i = 0] \tag{3}$$

We refer to a *disparate-impact-minimizing intervention policy* as a policy, Π, that provides a balanced relative-factual-outcome distribution with respect to a protected attribute S. Specifically, the following must hold for the outcome distribution imposed by Π and its corresponding W:

$$P(Y = 1 | W, S = 0) = P(Y = 1 | W, S = 1) \tag{4}$$

3 RELATED WORK

Most approaches to fair intervention policies can be categorized into two groups: preprocessing-based solutions and postprocessing-based solutions. Preprocessing-based solutions are mostly based on debiasing techniques [16, 29]. While debiasing techniques can be efficient they also have their limitations: the inability to adequately account for dependencies between the protected attributes and other variables [10]. Postprocessing works frame the problem as a constrained optimization problem, adding fairness constraints to the post-inference optimization problem [3, 19]. Such approaches become less efficient as the number of features increases.

Our research builds on existing works on algorithmic personalization [5, 6, 20, 24–26, 28, 30, 31, 37, 41, 43], some of which are explicitly aimed at introducing new uplift modeling techniques. However, such works have generally ignored any outcome-based fairness considerations in policy design, as well as fairness of allocation in dynamic intervention policies. Our work also builds on existing literature on bias and fairness [2, 4, 10, 11, 17, 18, 21, 32, 36, 38, 42] which focuses on uncovering biases and their causes as well as offering conceptual solutions for addressing such biases. Our work differs from the above works in that it offers a systematic approach to disparate-impact minimization which is both self-contained and thus does not require additional pre-processing or post-processing steps, and at the same time accounts for *all* sources of disparate impact including explicitly accounting for decisional disparities. Furthermore, unlike the above works, our work considers the notion of long-term disparate impact and presents models for balancing utility and outcome fairness in dynamic environments.

4 DISPARATE-IMPACT-MINIMIZING NETWORKS

In this section, we present our first approach for creating uplift models for disparate-impact-minimizing intervention policies: Disparate-Impact-Minimizing Networks (DIMN). A DIMN is a multi-task neural network constrained to minimize the disparate impact in the network's predictions. Here, we consider a *static* setting in which given a cost, C, all C individuals to be treated are chosen at the same point in time.

The network architecture is based on the idea of multi-task learning [12] in which multiple tasks are being learned in parallel by different parts of the same network thus on the one hand leveraging the ability to cross-learn among the shared parts of different tasks, gaining insights from one task and applying them for learning other similar tasks, and on the other hand allowing enough flexibility for each task to learn its own decision boundary.

Specifically, our network is composed of two types of layers: shared layers, L_s, and idiosyncratic layers, L_p. The shared layers are shared among all the tasks to be learnt. The idiosyncratic layers are task-specific. Our network incorporates $M + 1$ learning tasks, where M denotes the number of interventions that can be used. For simplicity of explanation, we assume now that $M = 2$. The first two separate, though related learning tasks are learning the response surface $E[Y_i^1 | X_i = x]$ and learning the response surface $E[Y_i^0 | X_i = x]$. Learning those two tasks using a multitask network has multiple benefits: on the one hand, the idiosyncratic layers can be used to compensate for the lack of flexibility of S-type models. Flexibility is an extremely important property in that context, especially in high-dimensional spaces where the response surfaces might have very different properties. On the other hand, having only a single model to estimate all outcomes, unlike other meta-models such

as the T-model, results in higher statistical accuracy and can also better account for imbalanced datasets [27, 31].

A key novel architectural feature of our model is what we refer to as "IPM regularization". Our model is regularized to reduce the disparate impact resulting from our interventions. This is done by adding a third (or more generally, $M+1$th) task to our multi-task network: an additional head aimed at regularizing our predictions using an Integral Probability Metric (IPM). An IPM measures the distance between two distributions; in our case, those distributions are $P(\hat{Y}=1|W,S=0)$ and $P(\hat{Y}=1|W,S=1)$. The loss associated with the regularization head aims at minimizing the IPM distance between the two distributions, thus constraining the predictions of $\hat{Y}^j$'s heads, used to compute $\hat{\tau}$.

We train the network in an end-to-end manner, jointly learning all the potential outcomes by minimizing both the regression losses and the IPM distance between the protected and unprotected distributions induced by the representation. This can be viewed as learning the functions m_0 and m_1 under a constraint that encourages a balanced positive outcome across populations with different values of S. The dataset, D, is seen as comprising two batches: a treated batch: customers i such that $W_i = 1$ and a control batch: customers i such that $W_i = 0$. The network is trained by alternating between the different "batches": at epoch e, we use all the batches to train all the shared layers, but only the idiosyncratic layers associated with batch $e\%M$. Our loss function is shown in Equation 5:

$$\mathcal{L}(Y,\hat{Y}) = \sum_{i=1}^{n}(1-W_i)*(\hat{Y}_i^0 - Y_i)^2 + W_i*(\hat{Y}_i^1 - Y_i)^2 + \delta * \mathcal{IPM}(P(\hat{Y}=1|W,S=0),(P(\hat{Y}=1|W,S=1))) \quad (5)$$

Where $\mathcal{IPM}$ denotes a concrete IPM scheme such as the Maximum Mean Discrepancy [39].

The fairness-accuracy trade off is expressed via the hyperparameter δ. When δ is set to 0, no fairness constraints will be imposed. At higher values of δ, the disparate impact resulting from our policy will decrease, yet beyond a certain point the policy's accuracy will also start decreasing. An interesting point to consider is that, beyond a certain value of δ, the decreasing policy's accuracy will also lead to an increasing disparate impact since, unlike disparate-treatment minimization, disparate-impact minimization necessitates the model to have a good accuracy so it can effectively predict the population's outcomes.

5 CUMULATIVE DISPARATE IMPACT

In the previous section, we considered a setting to which we referred as the *static* setting: given a cost, C, Π chooses all the C customers it treats at the same time. In this section, we consider an alternative setting: the dynamic setting. In the dynamic setting, featuring a *multi-step intervention policy*, a firm aims at making T sequential intervention decisions, $d_1 \dots d_T$, corresponding to T points in time, $t \in 1 \dots T$. Given such a T-step process, our goal is to design a dynamic uplift modeling technique that optimizes both the *cumulative* utility and the *cumulative* disparate impact resulting from the entire policy; that is, both maximizing the *global* policy's utility and minimizing its *global* disparate impact at time T instead of optimizing the *local* utility and disparate impact at the end of each intervention period, $t \in 1 \dots T$.

To better illustrate the problem, let us revisit the example given in Section 1. Assume that in January 2022, Amazon decides to formally launch its new Prime service in January 2023. Instead of a one-time promotional coupon campaign taking place in January 2022 and targeting C customers, Amazon decides to launch a monthly promotional coupon campaign with a dynamic targeting policy that may change every month throughout 2022, targeting $C\alpha$ ($0 < \alpha < 1$) customers every month. Having decided to launch a dynamic, monthly campaign, Amazon should also take a different approach to disparate-impact minimization. Since the promotional campaign ends in January 2023, Amazon must ensure that the *global* registered population — that is, the overall group of customers registered to the new service by January 2023 — is balanced ethnicity-wise. However, when determining each month's targeting policy, Π_t, Amazon can only observe the *local* race-based imbalance, γ_t, with respect to the current month, t, resulting from prior months' policies $\Pi_1 \dots \Pi_{t-1}$. Amazon's challenge is thus the following: for each targeting period, t, it should create a targeting policy that minimizes the *cumulative* disparate impact, $\mathcal{F}_T$, at the end of the T-step campaign (that is, in January 2023), given its current, local knowledge at time t of customer's features $X_{i,t}$; current race-based imbalance, γ_t; and the outcomes of prior targeting policies, $Y_{i,1} \dots Y_{i,t-1}$.

6 DISPARATE-IMPACT-MINIMIZING AGENTS

The goal of a Disparate-Impact-Minimizing Agent (DIMA) is designing fair multi-step intervention policies: a T-step intervention policy composed of T local decisions, aimed at maximizing the cumulative utility and minimizing the cumulative disparate impact. We implement Disparate-Impact-Minimizing Agents using a multi-task reinforcement learning framework, on which we elaborate in this section.

Notations We consider a similar setting to the one in prior sections where $W_{i,t}$ denotes whether customer i was treated at step t; $Y_{i,t}$ denotes the outcome of customer i at time t; $X_{i,t}$ denotes the feature vector of customer i at time t. We further introduce the notation $< t$, which denotes the statement "at each step before step t".

The building blocks of our framework are an *action*, a_t; a *state*, s_t; a reward, r_t; and two models: a CATE-prediction model, m^{pred}, and a policy-optimization model, m^{opt}. A key novel feature of our framework is its *multi-task optimization*, combining both CATE prediction and policy estimation. Our framework alternates between two steps: an optimization step and a prediction step, explicitly using the outcome of the prediction step at step t as features to the optimization step at time t, and implicitly using the outcome of the optimization step at time t to further enhance the prediction model used at the next prediction step at time $t+1$.

To illustrate the idea, let us revisit our coupon campaign example. Given the multi-step campaign Amazon launches from January 2022 to January 2023 ($T = 12$), let us assume we are now in March, designing the targeting policy of March's campaign. At the beginning of March, we are able to first observe February's registration decisions (outcomes) of all customers, $\{Y_{i,2}\}$. February's outcomes, as well as the corresponding feature values in February $\{X_{i,2}\}$ will be used for performing a prediction step, updating the CATE-prediction model, m_3^{pred}, thus resulting in more accurate predictions of $\hat{\tau}_i, \hat{Y}_i^0, \hat{Y}_i^1$. In the optimization step, we use the updated m_3^{pred} to predict the

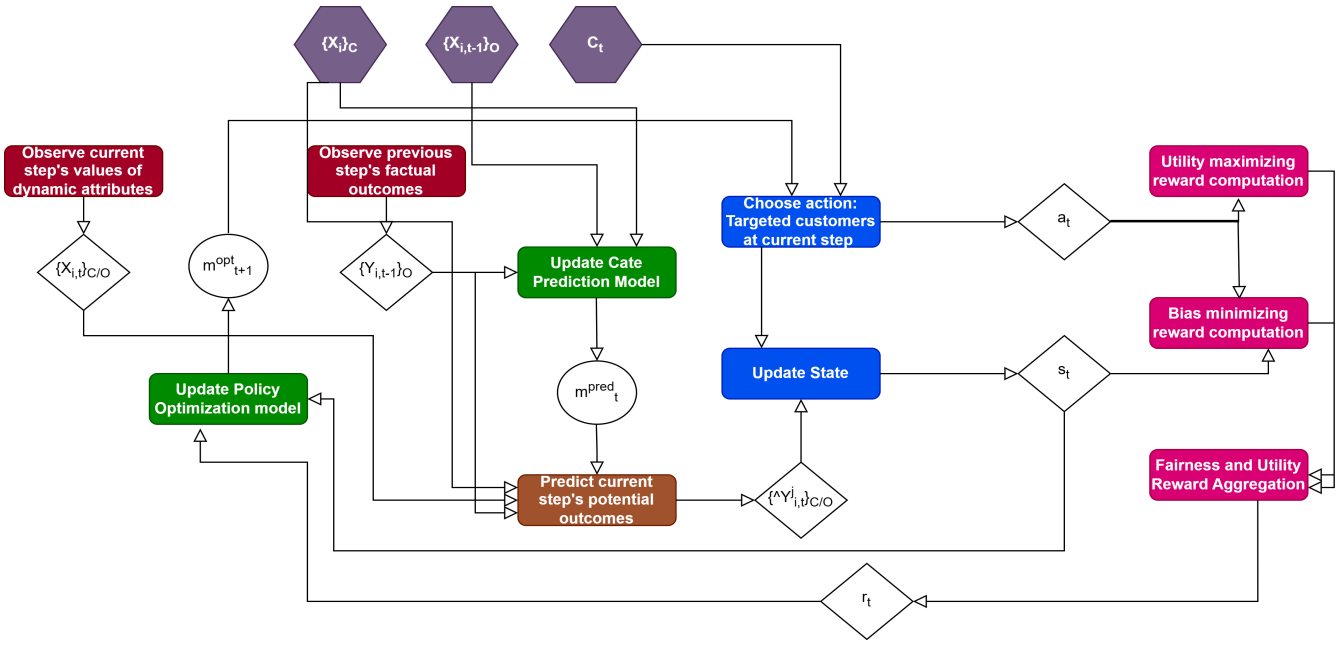


Figure 1: System model. $\{\}_C$ denotes the entire set of customers. $\{\}_O$ represents the subset of customers on which we have observed (*i.e.* factual) outcomes using their decisions at steps $1, \ldots, t-1$.

updated CATEs and potential outcomes with respect to March, $\hat{\tau}_3$, $\hat{Y}^0_{i,3}$, $\hat{Y}^1_{i,3}$. Those predictions, as well as other features as previously described, will be used as input for the policy-optimization model, m_3^{opt}; the output of m_3^{opt} corresponds to the agent's action in March: the subset of customers p_3 that will be targeted in March. The same process will repeat in April, where March's observed outcomes, $\{Y_{i,3}\}$, will be used to further incrementally train m_4^{pred} and create more accurate CATE predictions inputted to m_4^{opt}. This alternation between a prediction and an optimization step leads to faster convergence and more accurate solutions.

System model Our system model (step t) works as follows:

1. Prior outcome observation: The outcomes, $Y_{i,t-1}$, of all customers i such that $Y_{i,<t-1} = 0$ are observed.

2. Prediction-Model update: The CATE-prediction model, m_{t-1}^{pred}, is updated using the observed outcomes at the previous step, $Y_{i,t-1}$ and their corresponding features, $X_{i,t-1}$. This results in an updated prediction model, m_t^{pred}.

3. Updated CATE prediction: The agent uses m_t^{pred} to predict the current potential outcomes and CATEs of all customers i such that $Y_{i,<t} = 0$ and $W_{i,<t} = 0$.

4. State update: The predicted CATEs and the potential outcomes, $\hat{\tau_{i,t}}$, $\hat{Y^0_{i,t}}$, $\hat{Y^1_{i,t}}$, as well as the previous-step observed outcomes, $Y_{i,t-1}$, are used to update s_t.

5. Action selection: The policy-optimization model, m_t^{opt}, takes as input the current state s_t, and outputs an action: a subset of customers p_t, $|p_t| = C_t$ such that for each $i \in p_t$, $Y_{i,<t} = 0$ and $W_{i,<t} = 0$.

6. Reward assignment: The environment produces a scalar value, r_t. More details are given in Subsection 6.1.

7. State update: s_t is updated according to the consequences of the selected action, a_t. For instance, the bits in $[W^b_{i,t}]$, corresponding to customers in p_t are set; the system's state of imbalance, γ_t, is updated to reflect the new state of imbalance resulting from targeting the customers in p_t, *etc.*

Practically speaking, both m_t^{opt} and m_t^{pred} are implemented as neural networks. A neural-network-based policy-optimization model results in a deep-reinforcement-learning-based policy. m_t^{pred} is implemented as a DIMN without the IPM head.

6.1 Reward function

A utility-maximizing reward function Our first attempt at designing a utility-maximizing reward function is:

$$r_{utility} = \frac{\sum_{i \in p_t} \hat{\tau_{i,t}}}{\sum_{i=1}^{n} \hat{\tau_{i,t}}} \tag{6}$$

When using the reward function in Equation 6 we encountered two problems: first, the signal obtained by the reward function can become weak if either C_t is too small; n is too big; or the variance among customers' CATE is small. Second, considering only $\hat{\tau}$ for utility maximization did not seem to be enough for obtaining a high degree of utility for the binary-outcome setting and in datasets in which the correlation between τ and Y^1 is low. We thus revised our utility-maximizing sub-reward function (Equation 7):

$$\zeta(i) = \frac{\hat{\tau_{i,t}}}{1 - \xi * \hat{Y^1_{i,t}}} \quad r_{utility} = \frac{\sum_{i \in p_t} \zeta(i)}{\max\limits_{\substack{p_{max} \\ |p_{max}|=C_t}} \sum_{i \in p_{max}} \zeta(i)} \tag{7}$$

Where p_{max} denotes the subset of customers of size C_t with the largest cumulative ζ that could have been chosen at time t; that is, for each $i \in p_{max}$, both $W_{i,<t} = 0$ and $Y_{i,<t} = 0$.

The revised sub-reward contains two major changes compared to Equation 6: first, the denominator is now a much tighter upper bound on the highest sum of CATE increases we can obtain at step t, resulting in a meaningful signal that can adequately take the entire range of (0,1). Our second change is considering a function of Y^1, in addition to τ, when computing the reward. ζ will be high only if *both* τ and Y^1 are high, so as to prioritize customers that not only have a high CATE, but their baseline probability of a positive outcome is also high. The policymaker can control the weight that she assigns to Y^1 compared to τ using ξ.

A disparate-Impact-minimizing reward function A naive attempt of designing a disparate-impact-minimizing reward function is as follows:

$$r_{fairness} = \frac{\frac{\min(\sum_{i=1}^{n} \mathbb{1}_{\neg(Y_{i,<t}=0) \wedge S_i=0}, \sum_{i=1}^{n} \mathbb{1}_{\neg(Y_{i,<t}=0) \wedge S_i=1})}{\sum_{i=1}^{n} \mathbb{1}_{\neg(Y_{i,<t}=0)}}}{\frac{1}{2}} \tag{8}$$

Such a reward function will take its lowest value, 0, when only one group of customers — either that with $S = 0$ or with $S = 1$ — made positive-outcome decisions by time t. It will take its highest value, 1, where exactly half of the customers who made decisions leading to a positive outcome by time t are from the $S = 0$ group, and half are from the $S = 1$ group. This reward function aims at capturing the current state of positive-outcome-imbalance by considering the imbalance in positive outcome by the current step among the two demographic groups: customers with $S = 0$ and customers with $S = 1$. While such a reward function may seem like a good surrogate of our overall fairness objective, it has one critical flow: it only accounts for *past* positive outcomes, *i.e.* those at t', $t > t' > 0$, while failing to account for *future* positive outcomes, *i.e.* those at t', $T > t' > t$. Because our goal is minimizing the *cumulative* disparate impact, *i.e.* the disparate impact at the end of the T-step process, our reward function must account for *future* positive outcomes in addition to *past* positive outcomes. The reward function presented in Equation 9 aims at doing just that.

$$r_{fairness} = \frac{\frac{\min(\sum_{i=1}^{n} Y^*_{i,t}, \sum_{i=1}^{n} Y^*_{i,t})}{S_i=0 \qquad S_i=1}}{\frac{\sum_{i=1}^{n} Y^*_{i,t}}{\min(\sum_{i=1}^{n} \mathbb{1}_{S_i=0}, \sum_{i=1}^{n} \mathbb{1}_{S_i=1})}} \tag{9}$$
$$\frac{}{n}$$

$$Y^*_{i,t} = \begin{cases} 1 & \neg(Y_{i,<t} = 0) \\ \hat{Y}^1_{i,t} & \neg(W_{i,<t+1} = 0) \wedge Y_{i,<t} = 0 \\ \hat{Y}^0_{i,t} & W_{i,<t+1} = 0 \wedge Y_{i,<t} = 0 \end{cases} \tag{10}$$

Unlike Equation 8, which approximates the outcome imbalance at step t solely based on the subset of customers that made a positive-outcome decision by time t, Equation 9 approximates the outcome imbalance at time t based on *all* the customers, using the *predicted positive-outcome probabilities*, $Y^*_{i,t}$. Thus, the reward function in Equation 9 can be seen as a weighted version of the reward function in Equation 8, where weights correspond to the customer's predicted probability of making a positive-outcome decision at any step t', $t' => t$.

Table 1: Results — Canvassing dataset

Model	$\mathcal{U}$	$\mathcal{F}$
CR	.8	.27
CF	.77	.22
DIMN (No IPM)	.83	.19
DIMN (With IPM)	.83	.04

Table 2: Results — Campaign dataset

Model	$\mathcal{U}$	$\mathcal{F}$
CR	.59	.25
CF	.54	.27
DIMN (No IPM)	.6	.26
DIMN (With IPM)	.59	.06

Our reward function, R, is a weighted function of the two sub-reward functions:

$$R = \delta_1 * r_{utility} + \delta_2 * r_{fairness} \; 0 <= \delta_1, \delta_2 <= 1$$

Using δ_1, δ_2 the policymaker can explicitly control the utility-fairness tradeoff.

7 EXPERIMENTAL EVALUATION

To evaluate our Disparate-Impact-Minimizing Networks we use two publicly-available datasets: Marketing Campaign dataset [4] and Door-to-Door Canvassing dataset [8, 31].

Table 1 and Table 2 present our results on both datasets and compare them to three uplift-modeling baselines: a causal regression (CR) baseline and a causal-forest (CF) baseline [41], as well as a causal neural network baseline: a DIMN without its IPM head (implemented by setting the IPM's weight to 0). Utility ($\mathcal{U}$) is measured as the proportion of participants choosing brand's A product (Campaign dataset) and the proportion of individuals with positive Feeling Thermometer [33] towards transgender people (Canvassing dataset). Outcome-Imbalance ($\mathcal{F}$) is measured as the distance between proportions of different values of the protected attribute among individuals with a positive outcome. As our protected attribute, we use either the IsFemale attribute (Campaign dataset) or the AgeGroup attribute (Canvassing dataset). We randomly split each dataset into train and test samples and compute the inverse probability score (IPS) [23] estimator of the outcome generated by each allocation policy.

As can be seen from Table 1 and Table 2, the causal neural network's Utility results, having no fairness constraints, are either on par with or better than the other two baselines. DIMN's Utility results are on par with those of all baselines. Furthermore, DIMN's Outcome-Imbalance results are significantly better than those of all baselines. Notably, on the Canvassing dataset *DIMN achieved*

Figure 2: Results — Dynamic Interventions, $n = 5000$

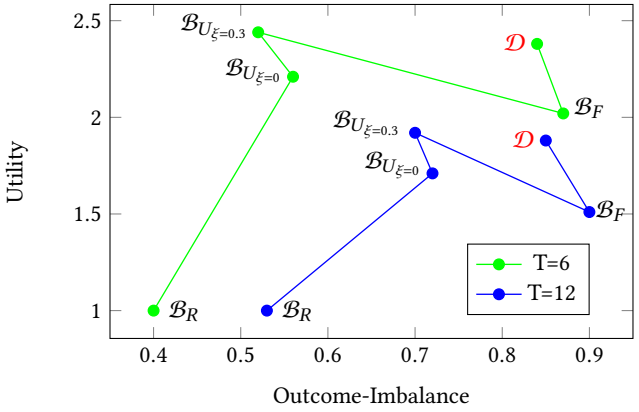
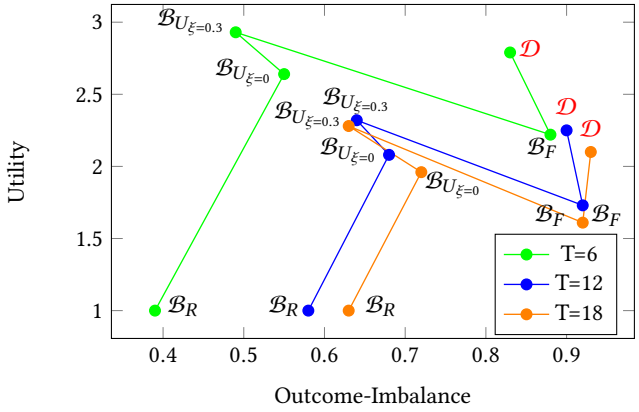


Figure 3: Results — Dynamic Interventions, $n = 20,000$



the same Utility results as the causal neural network — the "unfair" version of DIMN where $\delta = 0$. The results illustrate how DIMN is able to minimize disparate impact in intervention policies: We were able to design an *alternative feasible policy*, feasible as DIMN achieves similar Utility results as other, unfair policies, *with less adverse impact on the minority group*. To evaluate our Disparate-Impact-Minimizing agents we use a simulation-based approach. X_i is composed of four static, unprotected attributes:

$$X_i^d \sim Normal(0,1)\ d \in \{0,\ldots,2\}\ X_i^3 \sim Bernoulli(0.6) \tag{11}$$

Two dynamic, unprotected attributes, simulated as follows:

$$X_{i,t}^4 \sim Normal(0,1)\ X_{i,t}^5 \sim Bernoulli(0.3) \tag{12}$$

And one protected, static attribute, simulated as follows:

$$X_i^6 \sim Bernoulli(f(X_i^2)) \tag{13}$$

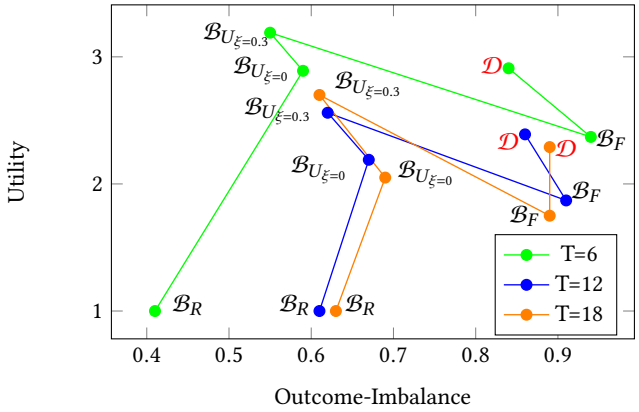
We simulate treatment effect using the following process:

$$\tau_{i,t} = \sum_{j=0}^{6} \alpha_j X_{i,t}^j + \sum_{j=0}^{6} \sum_{\substack{k=0 \\ k\neq j}}^{6} \beta_{j,k}(X_{i,t}^j * X_{i,t}^k)\ \alpha_j, \beta_{j,k} >= 0 \tag{14}$$

Figures 2-4 present the results of our agent ($\mathcal{D}$, $\delta_1, \delta_2 = 0.5$), implemented using Stable Baselines [22], compared to a random baseline ($\mathcal{B}_R$), an All-Fairness baseline ($\mathcal{B}_F$), and two All-Utility baselines ($\mathcal{B}_U$). An All-Fairness baseline is obtained by setting δ_1 to 0. An All-Utility baseline is obtained by setting δ_2 to 0; we present two All-Utility baselines: one where ξ is set to 0 and thus Y^1 is not taken into account in reward calculation, and one where ξ is set to 0.3.

Utility is measured as the proportion of customers with a positive outcome at the end of step T. To account for the fact that some customers make a positive-outcome decision independently of the chosen policy, Π, we compute Π's Utility as the ratio of the number of positive-outcome customers at step T when applying Π, to the number of positive-outcome customers at step T when applying $\mathcal{B}_R$. Outcome-Imbalance is measured as $1-\gamma_T$, where $0 <= \gamma_T <= 1$ denotes the proportion of minority-class customers (with respect to the protected attribute, S) with a positive outcome at the end of step T. As can be seen, our agent significantly outperforms the random

Figure 4: Results — Dynamic Interventions, $n = 50,000$



baseline on both the Utility and Outcome-Imbalance results. Furthermore, the All-Fairness DIMA baseline reaches an almost perfect Outcome-Imbalance score. The DIMA model adequately balances fairness and utility, achieving both high Utility and low Outcome-Imbalance. Notably, similar results are observed for both the case where $n = 5000$ and the case where $n = 50,000$, demonstrating the robustness of our approach. Finally, based on our All-Utility DIMA baselines it can be seen that considering Y^1 when calculating the episode's reward, in addition to τ yields higher results compared to a reward function that does not utilize Y^1.

8 BROADER IMPACT

This work sheds new light on the role of decisional disparities in technical modeling approaches aimed at reducing algorithmic bias. Accounting for decisional disparities in Uplift Modeling is essential for preventing firms from manipulatively misusing implicit correlations between the nature of the intervention and subjects' self-selected preferences to create seemingly neutral intervention policies that systematically exclude socially disadvantaged groups.

REFERENCES

[1] Access-Development. 2021. The Impact of Proximity on Consumer Purchases. https://shorturl.at/kCEFW.

[2] Krishna Acharya, Eshwar Ram Arunachaleswaran, Sampath Kannan, Aaron Roth, and Juba Ziani. 2023. Wealth dynamics over generations: Analysis and interventions. In *2023 IEEE Conference on Secure and Trustworthy Machine Learning (SaTML)*. IEEE, 42–57.

[3] Alekh Agarwal et al. 2018. A reductions approach to fair classification. In *ICML*.

[4] Eva Ascarza and Ayelet Israeli. 2022. Eliminating unintended bias in personalized policies using bias-eliminating adapted trees (BEAT). *PNAS* 119, 11 (2022).

[5] Susan Athey and Guido Imbens. 2016. Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences* 113, 27 (2016), 7353–7360.

[6] Susan Athey and Stefan Wager. 2021. Policy learning with observational data. *Econometrica* (2021).

[7] Gary S Becker. 2010. *The economics of discrimination*. University of Chicago press.

[8] David Broockman and Joshua Kalla. 2016. Durably reducing transphobia: A field experiment on door-to-door canvassing. *Science* (2016).

[9] Brookings. 2023. What the lack of premium grocery stores says about disinvestment in Black neighborhoods. https://shorturl.at/lAN37.

[10] Alessandro Castelnovo et al. 2021. The zoo of fairness metrics in machine learning. (2021).

[11] Sung-Ho Cho, Kei Kimura, Kiki Liu, Kwei-guu Liu, Zhengjie Liu, Zhaohong Sun, Kentaro Yahiro, and Makoto Yokoo. 2024. Fairness and efficiency trade-off in two-sided matching. In *AAMAS*.

[12] Ronan Collobert and Jason Weston. 2008. A unified architecture for natural language processing: Deep neural networks with multitask learning. In *ICML*. 160–167.

[13] DOH. 1968. HOUSING DISCRIMINATION UNDER THE FAIR HOUSING ACT. https://www.hud.gov/program_offices/fair_housing_equal_opp/fair_housing_act_overvie.

[14] DOL. 2024. Earnings Disparities by Race and Ethnicity. https://www.dol.gov/agencies/ofccp/about/data/earnings/race-and-ethnicity.

[15] Cynthia Dwork et al. 2012. Fairness through awareness. In *Proceedings of the 3rd innovations in theoretical computer science conference*.

[16] Michael Feldman et al. 2015. Certifying and removing disparate impact. In *SIGKDD*.

[17] Juexiao Feng, Yuhong Yang, Yanchun Xie, Yaqian Li, Yandong Guo, Yuchen Guo, Yuwei He, Liuyu Xiang, and Guiguang Ding. 2024. Debiased Novel Category Discovering and Localization. In *AAAI*, Vol. 38.

[18] Shaz Furniturewala, Surgan Jandial, Abhinav Java, Simra Shahid, Pragyan Banerjee, Balaji Krishnamurthy, Sumit Bhatia, and Kokil Jaidka. 2024. Evaluating the Efficacy of Prompting Techniques for Debiasing Language Model Outputs. In *AAAI*, Vol. 38.

[19] Gabriel Goh. 2016. Satisfying real-world goals with dataset constraints. In *NeurIPS*.

[20] P Richard Hahn et al. 2020. Bayesian regression tree models for causal inference. *Bayesian Analysis* 15, 3 (2020), 965–1056.

[21] Hoda Heidari and Jon Kleinberg. 2021. Allocating opportunities in a dynamic model of intergenerational mobility. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. 15–25.

[22] Ashley Hill et al. 2018. Stable Baselines. https://github.com/hill-a/stable-baselines.

[23] Daniel G Horvitz and Donovan J Thompson. 1952. A generalization of sampling without replacement from a finite universe. *Journal of the American statistical Association* (1952).

[24] Lihi Idan. 2022. Beyond purchase intentions: Mining behavioral intentions of social-network users. *International Journal of Human-Computer Interaction* 40, 5 (2022), 1111–1132. https://doi.org/10.1080/10447318.2022.2132195

[25] Lihi Idan. 2022. A Network-Based, Multidisciplinary Approach to Intention Inference. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*. 1–7. https://doi.org/10.1145/3491101.3519754

[26] Lihi Idan. 2022. Temporal-Attribute Inference Using Dynamic Bayesian Networks. In *International Conference on Computational Science*. Springer, 638–652. https://doi.org/10.1007/978-3-031-08754-7_67

[27] Lihi Idan. 2024. Towards Unsupervised Validation of Anomaly Detection Models. In *27th European Conference on Artififcial Intelligence (ECAI)*. https://doi.org/10.3233/FAIA240859

[28] Lihi Idan and Joan Feigenbaum. 2019. Show me your friends, and I will tell you whom you vote for: Predicting voting behavior in social networks. In *Proceedings of the 2019 IEEE/ACM international conference on advances in social networks analysis and mining*. 816–824. https://doi.org/10.1145/3341161.3343676

[29] James E Johndrow and Kristian Lum. 2019. An algorithm for removing sensitive information. *The Annals of Applied Statistics* 13, 1 (2019), 189–220.

[30] Jon Kleinberg. 2024. Revisiting the Behavioral Foundations of User Modeling Algorithms. In *Proceedings of the ACM on Web Conference 2024*. 1–1.

[31] Sören R Künzel et al. 2019. Metalearners for estimating heterogeneous treatment effects using machine learning. *PNAS* 116, 10 (2019), 4156–4165.

[32] Hussein Mouzannar, Mesrob I Ohannessian, and Nathan Srebro. 2019. From fair decision making to social equality. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*. 359–368.

[33] Aaron T Norton and Gregory M Herek. 2013. Heterosexuals' attitudes toward transgender people: Findings from a national probability sample of US adults. *Sex roles* 68 (2013), 738–753.

[34] Pew-Research-Center. 2020. Online Shopping and E-Commerce. https://shorturl.at/tVZ08/.

[35] Pew-Research-Center. 2022. For shopping, phones are common and influencers have become a factor – especially for young adults. https://www.pewresearch.org/short-reads/2022/11/21/for-shopping-phones-are-common-and-influencers-have-become-a-factor-especially-for-young-adults/.

[36] Devin G Pope and Justin R Sydnor. 2011. Implementing anti-discrimination policies in statistical profiling models. *American Economic Journal: Economic Policy* 3, 3 (2011).

[37] Soroush Saghafian and Lihi Idan. 2024. Effective generative AI: The human-algorithm centaur. *arXiv preprint arXiv:2406.10942* (2024).

[38] Aditya Shinde and Prashant Doshi. 2024. Modeling Cognitive Biases in Decision-Theoretic Planning for Active Cyber Deception. In *AAMAS*.

[39] Alex Smola et al. 2007. A Hilbert space embedding for distributions. In *International conference on algorithmic learning theory*.

[40] SOC-Investment-Group. 2024. Amazon. https://shorturl.at/prwDS.

[41] Stefan Wager and Susan Athey. 2018. Estimation and inference of heterogeneous treatment effects using random forests. *J. Amer. Statist. Assoc.* (2018).

[42] Cheng Yang, Jixi Liu, Yunhe Yan, and Chuan Shi. 2024. FairSIN: Achieving Fairness in Graph Neural Networks through Sensitive Information Neutralization. In *AAAI*, Vol. 38.

[43] Hema Yoganarasimhan et al. 2020. Design and evaluation of personalized free trials. *arXiv:2006.13420* (2020).