**AI and Data Science Journal**
**Online ISSN: 3078-7394**
**DOI: https://doi.org/10.61784/adsj4044**

# A SCALABLE CAUSAL MACHINE LEARNING FRAMEWORK FOR DATA-DRIVEN DECISION OPTIMIZATION

KaiLang Wei[1*], Zhou Lu[2], XuanPing Zou[3]
[1]*Emory University, Atlanta 30322, USA.*
[2]*Northeastern University, San Jose 95113, USA.*
[3]*Zoom, Inc., San Jose 95113, USA.*
[*]*Corresponding Author: KaiLang Wei*

**Abstract:** Allocating costly interventions such as incentives, outreach messages or preventive programs to the individuals who benefit most is a central task of data-driven decision making. It requires counterfactual predictions for several competing treatments, a principled way to choose among the many available causal learners, and an optimizer that respects budgets at the scale of millions of units. This paper presents SCOPE (Scalable Causal Orthogonal Policy Ensemble), an end-to-end framework that couples orthogonal machine learning with budgeted multi-treatment optimization. SCOPE builds a cross-fitted nuisance cache with histogram gradient boosting, trains a leakage-free library of S-, T-, X-, R- and DR-learners, aggregates them by arm-wise orthogonal stacking with non-negative, free-scale weights fitted on a doubly robust loss, and solves the resulting multiple-choice knapsack problem with a dual-price bisection whose gap to the linear relaxation is bounded by the gain of a single unit. Every stage runs in time linear in the sample size. On 50 IHDP replications SCOPE attains the lowest √PEHE (2.12 versus 2.14) and the highest policy value at every budget. On a field experiment with 305,866 voters and four arms it raises the turnout gain of a 10% mailing budget by 8.9% over the best single learner and by 17.8% over random targeting, and on a confounded multi-treatment benchmark it reaches 92.4–95.2% of the oracle value. The dual allocator matches the linear-programming relaxation within 0.010% while running about 83 times faster.
**Keywords:** Causal machine learning; Heterogeneous treatment effects; Uplift modeling; Doubly robust learning; Stacking; Budget-constrained optimization; Policy learning; Scalability

## 1 INTRODUCTION

Organizations increasingly decide whom to target with costly interventions: platforms choose which customers receive which incentive, campaigns decide which citizens to contact, and health systems select patients for preventive programs. These decisions are causal, because the relevant quantity is the change in an outcome that an intervention would produce rather than the outcome itself, and they are constrained by budgets. Causal machine learning supplies the required estimators, from meta-learners and double/debiased machine learning to causal forests and representation learning [1-9].

Turning these estimators into reliable decisions at scale remains difficult for three reasons. First, no conditional average treatment effect (CATE) estimator dominates across data-generating processes, and model selection is complicated by the absence of counterfactual labels [10]. Accuracy in estimating effects is, moreover, not the same as the quality of the resulting decisions: in our IHDP experiments the S-learner has one of the largest estimation errors yet ranks among the best single learners for targeting [11]. Second, realistic problems involve several treatments with heterogeneous costs under a global budget, which yields a multiple-choice knapsack problem (MCKP) over millions of units; industrial systems address it with bespoke estimators and solvers [12-15]. Third, cross-fitting, which is essential for orthogonality, multiplies the computational burden of every learner, and the cost of flexible estimators grows quickly on commodity hardware.

This paper proposes SCOPE, a scalable causal machine learning framework for data-driven decision optimization. SCOPE organizes learning and optimization in four linear-time stages: a cross-fitted nuisance cache, a leakage-free library of meta-learners, orthogonal arm-wise stacking, and a dual-price allocator. The main contributions are as follows.

1) An end-to-end framework that unifies orthogonal nuisance learning, heterogeneous meta-learners, stacking and budgeted allocation over multiple treatments with heterogeneous costs, built on a single histogram gradient boosting engine with linear cost in every stage.

2) Arm-wise orthogonal stacking with non-negative, free-scale weights and arm-specific intercepts fitted on a doubly robust loss. We show why calibrating the scale and level of each arm matters for cost-aware decisions, and we identify and remove a cross-fitting leakage that biases stacking toward learners that share nuisance estimates with the pseudo-outcomes.

3) An analysis that links the stages: second-order sensitivity of the stacking loss to nuisance errors, a regret bound that translates arm-wise CATE error into decision regret under budgets, and an $O(nK \log(1/\varepsilon))$ allocator whose gap to the linear relaxation is at most the gain of one unit.

© By the Author(s) 2024, under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0).

4) An evaluation on the IHDP benchmark, a 305,866-voter field experiment with four arms and a confounded multi-treatment simulator, including ablations and a scalability study, in which SCOPE delivers the highest policy value on all three problems.
Section 2 reviews related work, Section 3 formalizes the problem, Section 4 presents SCOPE and its analysis, Section 5 reports the experiments, and Section 6 concludes.

## 2 RELATED WORK

### 2.1 Heterogeneous Treatment Effect Estimation

Meta-learners reduce CATE estimation to supervised learning. The S-learner fits one outcome model with the treatment as a feature, the T-learner fits one model per arm, and the X-learner imputes individual effects and combines arm-specific regressions through the propensity score [1]. The R-learner minimizes a Robinson-type residual loss, and the DR-learner regresses doubly robust pseudo-outcomes and attains oracle rates under weak conditions [2-3]. Both are instances of orthogonal statistical learning, in which first-stage errors affect the excess risk only at second order [4-5]. Generalized random forests, Bayesian additive regression trees and neural architectures provide further function classes, uplift modeling studies the same task in marketing, and conformal meta-learners add predictive inference for individual effects [6-9, 16-18].

### 2.2 Policy Learning and Decision-Focused Learning

Policy learning optimizes treatment rules directly. Empirical welfare maximization, doubly robust policy learning and its multi-action extension provide regret guarantees for restricted policy classes [19-22]. In operations research, predict-then-optimize and decision-focused learning integrate estimation with the downstream optimization problem, and Fernández-Loría and Provost argue that causal decision making calls for criteria different from those of causal effect estimation [11, 23-25]. SCOPE keeps the plug-in structure, which leaves the optimizer exact and fast for knapsack-type constraints, and makes the effect scores decision-ready through stacking and arm-wise calibration.

### 2.3 Budget-Constrained Treatment Allocation

Large platforms allocate incentives under budgets. Zhao and Harinen combine multi-treatment uplift models with cost optimization, LBCF couples a unified causal forest with a dual gradient bisection solver, and DHCL learns a decision factor that bridges machine learning and operations research [13-15]. These systems are tied to particular estimators, whereas SCOPE provides an estimator-agnostic layer that aggregates a library before optimization, with an allocator grounded in the linear-programming theory of the MCKP [12].

### 2.4 Ensembles and Model Selection

Stacking combines base learners through cross-validated predictions; Breiman recommended non-negative weights, and the super learner enjoys oracle guarantees [26-27]. For causal effects the target is unobserved, so selection relies on surrogate losses whose behavior depends on the data-generating process and on the congeniality between criteria and estimators [10]. SCOPE stacks on the doubly robust pseudo-outcome loss, guarantees that out-of-fold predictions are independent of the evaluation targets, and estimates weights separately for every treatment arm.

## 3 PROBLEM FORMULATION

Consider $n$ units with covariates $X_i \in \mathbb{R}^d$, a treatment $T_i \in \{0, 1, \ldots, K\}$ in which 0 denotes control, and an outcome $Y_i$. In the potential outcomes framework $Y_i(k)$ is the outcome under treatment $k$ and $Y_i = Y_i(T_i)$. Let $\mu_k(x) = E[Y(k) \mid X = x]$, $e_k(x) = P(T = k \mid X = x)$ and $\tau_k(x) = \mu_k(x) - \mu_0(x)$. Under consistency, unconfoundedness and overlap, $e_k(x) \geq \eta > 0$, the effects $\tau_k$ are identified from the observed data.
Assigning treatment $k$ to unit $i$ costs $c_k(X_i) \geq 0$ with $c_0 \equiv 0$, and the total budget is $B$. The decision maker chooses $a_i \in \{0, \ldots, K\}$ to solve

$$\max_a \Sigma_i \tau_{ai}(X_i) \qquad s.t. \qquad \Sigma_i c_{ai}(X_i) \leq B, \tag{1}$$

with $\tau_0 \equiv 0$. Problem (1) is an MCKP in which each unit is a class of $K + 1$ items [12]. We measure the per-capita policy value $V(a) = n^{-1} \Sigma_i \tau_{ai}(X_i)$ and the regret $V(a^*) - V(a)$ relative to the oracle allocation $a^*$ that solves (1) with the true effects. Because $\tau_k$ is unknown, SCOPE learns decision scores $\tilde{\tau}_k$ and solves (1) with $\tilde{\tau}_k$ in place of $\tau_k$.

## 4 THE SCOPE FRAMEWORK

### 4.1 Overview

Figure 1 summarizes the four stages. All learning is cross-fitted over $F$ folds and shares one histogram gradient boosting decision tree (GBDT) engine, whose training cost grows linearly with the sample size [28]. Algorithm 1 lists the complete procedure.

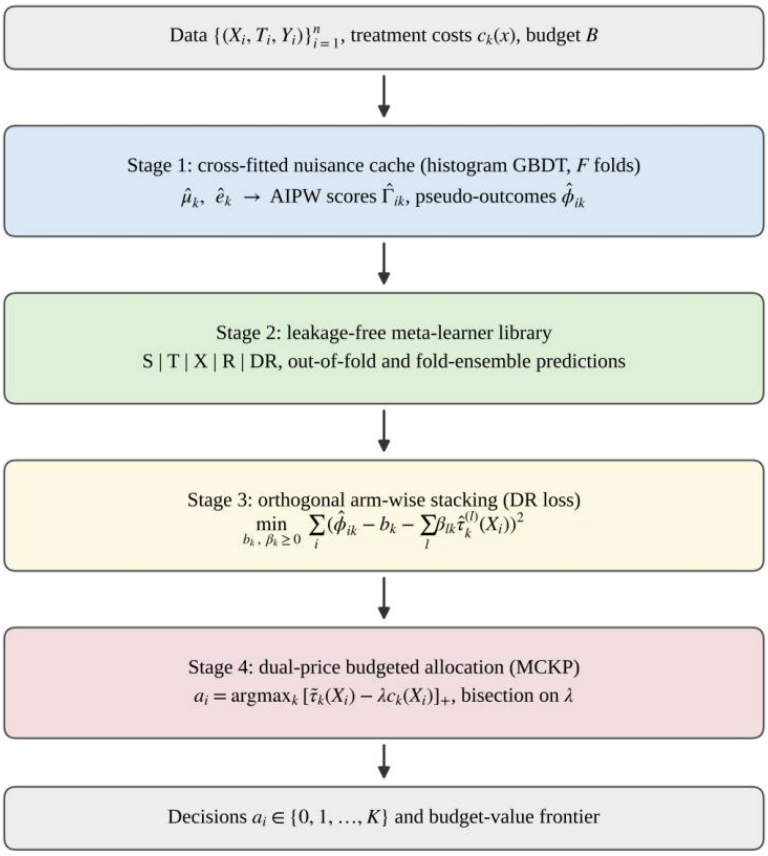



**Figure 1** Overview of SCOPE

Stages 1–3 are cross-fitted over $F$ folds on a single histogram GBDT engine; Stage 4 solves the budgeted multi-treatment allocation by dual-price bisection.

### 4.2 Stage 1: Cross-Fitted Nuisance Cache

The data are split into $F$ folds stratified by treatment. For every fold $f$, arm-specific outcome models $\hat{\mu}_k^{(-f)}$ and a multinomial propensity model $\hat{e}^{(-f)}$ are fitted on the remaining folds; known assignment probabilities are used instead in randomized experiments. The out-of-fold predictions define the augmented inverse propensity weighted (AIPW) scores

$$\hat{\Gamma}_{ik} = \hat{\mu}_k(X_i) + 1\{T_i = k\}(Y_i - \hat{\mu}_k(X_i)) / \hat{e}_k(X_i), \tag{2}$$

and the effect pseudo-outcomes $\hat{\varphi}_{ik} = \hat{\Gamma}_{ik} - \hat{\Gamma}_{i0}$ for $k = 1, \ldots, K$. With exact nuisances $E[\varphi_{ik} \mid X_i] = \tau_k(X_i)$, and the same scores support doubly robust policy evaluation [20]. Propensities are clipped at 0.03 and renormalized.

### 4.3 Stage 2: Leakage-Free Meta-Learner Library

The library $\mathscr{L}$ contains five learners with different inductive biases: S (one GBDT on the covariates and a one-hot treatment code), T (difference of arm-specific regressions), X (propensity-weighted imputed-effect regressions), R (residual-on-residual regression on the arms $\{0, k\}$) and DR (regression of $\hat{\varphi}_{ik}$ on $X_i$) [1-3]. For every fold $f$, each learner is trained on the other folds and predicts fold $f$, which yields out-of-fold predictions $\hat{\tau}_k^{(l)}(X_i)$; at deployment the $F$ fold models are averaged into a cross-fit ensemble $\bar{\tau}_k^{(l)}$.

A subtle problem arises when library members are trained on targets taken from the cache. For a unit $j$ outside fold $f$, the cached prediction $\hat{\mu}_k(X_j)$ was fitted on data that include fold $f$. A learner trained on such targets therefore depends on the outcomes of fold $f$, its out-of-fold prediction becomes correlated with the noise in $\hat{\varphi}_{ik}$, and the stacking loss underestimates its risk, a mechanism related to the congeniality bias discussed in [10]. With this shortcut our synthetic experiments assigned more than 90% of the stacking weight to the X-learner and lost about six points of policy value. SCOPE removes the dependence: within the training folds of $f$ the T- and X-learners use the fold-specific models $\mu^{(-f)}$, and the R- and DR-learners use an inner two-fold cross-fit. Every out-of-fold prediction for fold $f$ is thus a function of data outside fold $f$, while the pseudo-outcomes of fold $f$ come from models that never saw fold $f$.

### 4.4 Stage 3: Orthogonal Arm-Wise Stacking

For every arm $k$, SCOPE fits

$$(\hat{b}_k,\ \hat{\beta}_k) = \text{argmin}_{b \in \mathbb{R},\, \beta \geq 0} \Sigma_i\, (\hat{\varphi}_{ik} - b - \Sigma_l \beta_l\ \hat{\tau}_k^{(l)}(X_i))^2, \tag{3}$$

and scores units by $\tilde{\tau}_k(x) = \hat{b}_k + \Sigma_l \hat{\beta}_{lk} \bar{\tau}_k^{(l)}(x)$. Problem (3) is solved exactly by centering, reducing the $M \times M$ Gram matrix with a Cholesky factorization and applying non-negative least squares, at cost $O(nKM^2 + KM^3)$ for $M = |\mathscr{L}|$ learners. The weights are not forced to sum to one, so the stack can undo the shrinkage of individual learners, and the intercept anchors the mean of $\tilde{\tau}_k$ to the doubly robust estimate of the average effect of arm $k$.

*Proposition 1 (orthogonality of the stacking loss):* Let $\varphi_k$ be the pseudo-outcome built with the true nuisances and $r_k(x) = E[\hat{\varphi}_k - \varphi_k \mid X = x]$. For every candidate $g$, $E[(\hat{\varphi}_k - g)^2] = E[(\tau_k - g)^2] + 2E[r_k(\tau_k - g)] + C_k$, where $C_k$ does not depend on $g$ and $|r_k(x)| \le \Sigma_{j=0,k} |\hat{e}_j(x) - e_j(x)| \, |\hat{\mu}_j(x) - \mu_j(x)| / \hat{e}_j(x)$.

*Proof sketch:* Expand $\hat{\varphi}_k - g = (\hat{\varphi}_k - \varphi_k) + (\varphi_k - \tau_k) + (\tau_k - g)$; the cross term with $\varphi_k - \tau_k$ vanishes because $E[\varphi_k \mid X] = \tau_k$, and for each arm the AIPW bias equals $(\hat{\mu}_j - \mu_j)(\hat{e}_j - e_j)/\hat{e}_j$. ▌ Consequently, the stacking risk of every candidate differs from its true mean squared error by at most $2\|r_k\| \, \|\tau_k - g\|$, a product of nuisance errors, and the weights in (3) inherit an oracle inequality of the super-learner type [3, 5, 27].

*Proposition 2 (calibration matters across arms):* With one treatment and uniform costs, the allocation obtained from $\tilde{\tau}$ is invariant to every transformation $\alpha\tilde{\tau} + b$ with $\alpha > 0$ whenever the budget binds. With $K \ge 2$ arms or heterogeneous costs, the net scores $\tilde{\tau}_k - \lambda c_k$ are compared across arms, so arm-specific errors in scale or level change decisions even when every within-arm ranking is perfect. The first claim holds because a binding budget with uniform costs selects the top-ranked units; for the second, two units with identical rankings but rescaled arm scores already swap their optimal arms. Arm-wise intercepts and free-scale weights therefore target exactly the quantities that enter cross-arm decisions.

### 4.5 Stage 4: Dual-Price Budgeted Allocation

Relaxing the budget in (1) with a price $\lambda \ge 0$ separates the problem across units:

$$a_i(\lambda) = \operatorname{argmax}_k [\tilde{\tau}_k(X_i) - \lambda \, c_k(X_i)], \tag{4}$$

with ties broken toward control. The total cost $C(\lambda)$ is non-increasing in $\lambda$, so a bisection over $[0, \max_{ik} \tilde{\tau}_k/c_k]$ finds the smallest feasible price $\hat{\lambda}$ in $O(\log(1/\varepsilon))$ iterations of $O(nK)$ work, after which the residual budget is spent on the upgrades with the largest benefit-to-cost ratio. A dual bisection is also used by LBCF; SCOPE couples it with stacked, arm-calibrated scores and the residual fill [14].

*Proposition 3 (near-optimality):* Let $OPT_{LP}$ be the value of the linear relaxation of (1) under the scores $\tilde{\tau}$. For scores in general position and up to the bisection tolerance, the dual allocation $\hat{a}$ satisfies $\Sigma_i \tilde{\tau}_{\hat{a}i}(X_i) \ge OPT_{LP} - \delta$ with $\delta = \max_{i,k} \tilde{\tau}_k(X_i)^+$. By linear-programming duality $OPT_{LP} = \min_\lambda \{\Sigma_i \max_k [\tilde{\tau}_k - \lambda c_k]^+ + \lambda B\}$, and at the minimizing price the relaxation has at most one fractional class; the bisection solution differs from it only in that class, and the residual fill never decreases the objective [12].

### 4.6 Decision Regret

*Theorem 1:* Let $a^*$ solve (1) under the true effects and let $\hat{a}$ be the dual allocation computed from $\tilde{\tau}$. Then

$$V(a^*) - V(\hat{a}) \le (2/n) \, \Sigma_i \max_k |\tilde{\tau}_k(X_i) - \tau_k(X_i)| + \delta/n. \tag{5}$$

*Proof:* Write $V_\tau$ and $V_{\tilde{\tau}}$ for the values under $\tau$ and $\tilde{\tau}$, and decompose $V_\tau(a^*) - V_\tau(\hat{a})$ into $[V_\tau(a^*) - V_{\tilde{\tau}}(a^*)] + [V_{\tilde{\tau}}(a^*) - V_{\tilde{\tau}}(\hat{a})] + [V_{\tilde{\tau}}(\hat{a}) - V_\tau(\hat{a})]$. The first and third terms are each bounded by the average arm-wise error, and the middle term is at most $\delta/n$ by Proposition 3 because $a^*$ is feasible for the relaxation. ▌ Theorem 1 shows that decision regret is controlled by the arm-wise CATE error that Stage 3 minimizes, and Proposition 2 explains why the calibration of each arm, not only the ranking within it, enters this error.

### 4.7 Complexity and Scalability

Let $T_{GB}(n) = O(ndJ)$ be the cost of a histogram GBDT with $J$ trees on $n$ samples. Stage 1 costs $O(F(K + 1)T_{GB}(n))$, Stage 2 costs $O(FKT_{GB}(n))$ including the inner cross-fit, Stage 3 costs $O(nKM^2)$ and Stage 4 costs $O(nK \log(1/\varepsilon))$, so the whole pipeline is linear in $n$ with memory $O(nKM)$. The fits within Stages 1 and 2 are independent jobs that can be distributed across cores or machines.

**Algorithm 1** SCOPE

**Input:** data $\{(X_i, T_i, Y_i)\}$, costs $c_k(\cdot)$, budget $B$, folds $F$, library $\mathscr{L}$
1: Split the data into $F$ folds stratified by $T$.
2: **for** $f = 1, \ldots, F$: fit $\hat{\mu}_k^{(-f)}$ and $\hat{e}^{(-f)}$ on the other folds; predict fold $f$.
3: Compute $\Gamma_{ik}$ and $\varphi_{ik}$ by (2).
4: **for** $f = 1, \ldots, F$ and $l \in \mathscr{L}$: train learner $l$ on the other folds with fold-specific nuisances; predict $\tau^{(l)}$ on fold $f$.
5: **for** $k = 1, \ldots, K$: solve (3) for $(\hat{b}_k, \hat{\beta}_k)$.
6: Score deployment units with $\tilde{\tau}_k(x) = \hat{b}_k + \Sigma_l \hat{\beta}_{lk} \bar{\tau}_k^{(l)}(x)$.
7: Bisect on $\lambda$ until $C(\lambda) \le B$, assign by (4) and fill the residual budget.
**Output:** allocation $\hat{a}$ and decision scores $\tilde{\tau}$.

## 5 EXPERIMENTS

### 5.1 Datasets and Protocol

*IHDP.* The Infant Health and Development Program benchmark contains 747 children with 25 covariates and a binary treatment; outcomes follow response surface B, so the true effects are known [16]. We use 50 replications with random 70/30 train/test splits.

*GOTV.* The social-pressure field experiment of Gerber et al. randomized 305,866 registered voters in Michigan to a control group (62.5%) or one of three mailings (12.5% each): civic duty, Hawthorne (notice of being studied) and neighbors (publication of household turnout) [29]. The outcome is turnout in the August 2006 primary, and the covariates are sex, age, turnout in the 2004 primary and household size. We train on a random 30% (91,759 voters) and evaluate on the remaining 214,107 voters with the design-based AIPW estimator, which uses the known assignment probabilities and outcome models fitted on the training split. The doubly robust average effects on the evaluation set are 1.85, 2.77 and 8.22 percentage points (pp) for the three mailings. Mailings have unit cost and the budget is a share of the electorate.

*Synthetic.* A marketing simulator with 20 covariates (10 Gaussian, 5 binary, 5 uniform), three incentives with nonlinear and partly negative effects, covariate-dependent costs (base costs 0.5, 1.0 and 1.5 scaled by a customer-specific factor in [0.75, 1.25]), confounded historical assignment with minimum propensities below 0.01, and heteroscedastic noise. We use 20,000 training and 20,000 test units for 3 seeds; the budget is expressed as a share of the cost of giving the cheapest incentive to everyone. The same generator supplies the samples of the scalability study (Table 1).

**Table 1** Datasets Used in the Experiments

| Dataset | Units | Covariates | Arms | Assignment | Evaluation |
|---|---|---|---|---|---|
| IHDP | 747 | 25 | 2 | Observational | True CATE |
| GOTV | 305,866 | 4 | 4 | Randomized | Design AIPW |
| Synthetic | 40,000 (≤100,000) | 20 | 4 | Confounded | True CATE |

### 5.2 Baselines and Implementation

We compare SCOPE with the S-, T-, X-, R- and DR-learners and with the causal forest with orthogonalized residuals (CausalForestDML in EconML), which implements generalized random forests [1-3, 7]. All learners share the same LightGBM base model with sample-size-dependent hyperparameters (for example 300 trees, 16 leaves and learning rate 0.05 for $3{,}000 \le n < 50{,}000$), and all effect estimates are passed to the same dual allocator, so differences reflect the decision quality of the scores [28]. SCOPE uses $F = 3$ folds and the library $\{S, T, X, R, DR\}$. For IHDP and the simulator we report the policy value as a percentage of the oracle value, $V(\hat{a})/V(a^*)$, together with $\sqrt{PEHE}$ and, for IHDP, the absolute error of the average treatment effect $\varepsilon_{ATE}$; for GOTV we report the per-capita turnout gain in pp. All runs use a single CPU core.

### 5.3 Results on IHDP

**Table 2** Results on IHDP (Mean ± Standard Deviation over 50 Replications)

| Method | √PEHE | $\varepsilon_{ATE}$ | V@10% | V@20% | V@30% | V@50% |
|---|---|---|---|---|---|---|
| S-learner | 3.45±5.97 | 0.33±0.74 | 97.1±2.2 | 97.1±1.8 | 97.5±1.8 | 97.7±1.9 |
| T-learner | 2.14±3.41 | **0.22±0.32** | 96.5±2.0 | 97.0±1.6 | 97.2±1.5 | 97.9±1.2 |
| X-learner | 3.04±4.64 | 0.29±0.46 | 94.5±2.9 | 94.9±2.4 | 95.5±2.1 | 96.3±1.7 |
| R-learner | 2.95±4.00 | 0.53±0.92 | 93.1±3.1 | 93.8±2.2 | 94.3±2.1 | 95.2±1.6 |
| DR-learner | 2.94±3.21 | 0.29±0.43 | 91.1±4.7 | 92.8±3.4 | 93.7±2.6 | 94.9±2.3 |
| Causal forest | 4.11±6.65 | 0.76±1.65 | 94.1±3.8 | 94.6±2.8 | 94.5±2.9 | 94.7±3.1 |
| **SCOPE (ours)** | **2.12±3.41** | 0.24±0.33 | **97.2±2.2** | **97.7±1.3** | **98.0±1.3** | **98.5±1.0** |

Table 2 and Figure 2 report the results. SCOPE achieves the highest policy value at every budget, reaching 97.21%, 97.72%, 98.03% and 98.47% of the oracle value at budgets of 10%, 20%, 30% and 50%, and it also attains the lowest √PEHE (2.12). Its standard deviation across replications is the smallest at budgets of 20% and above, which indicates that stacking stabilizes the decisions as well as improving them on average. One-sided Wilcoxon signed-rank tests over the replications reject equality in favor of SCOPE ($p < 0.05$) in 23 of the 24 method–budget comparisons; the remaining cases are S-learner at 10%.

Figure 2(b) illustrates the gap between estimation and decision quality [11]. The S-learner has a large √PEHE (3.45) because its regularized trees shrink effect magnitudes, yet it is the strongest single learner at the 10% budget because its ranking of units is accurate. The DR-learner shows the opposite pattern. SCOPE assigns most of its weight to the S- and T-learners (average weights 0.40 and 0.39) and combines the ranking quality of the former with the calibration of the latter.

### 5.4 Results on the GOTV Field Experiment

**Table 3** Turnout Gain on the GOTV Evaluation Set (pp per Voter)

| Method | V@10% | V@20% | V@30% | V@50% |
|---|---|---|---|---|
| Random targeting | 0.822 | 1.644 | 2.466 | 4.110 |
| S-learner | 0.889 | 1.829 | 2.610 | 3.953 |

| Method | V@10% | V@20% | V@30% | V@50% |
|---|---|---|---|---|
| T-learner | 0.782 | 1.467 | 2.162 | 3.464 |
| X-learner | 0.783 | 1.577 | 2.379 | 3.412 |
| R-learner | 0.647 | 1.451 | 1.931 | 3.164 |
| DR-learner | 0.521 | 1.425 | 1.842 | 3.219 |
| Causal forest | 0.791 | 1.576 | 2.218 | 3.271 |
| **SCOPE (ours)** | **0.968** | **1.873** | **2.760** | **4.412** |

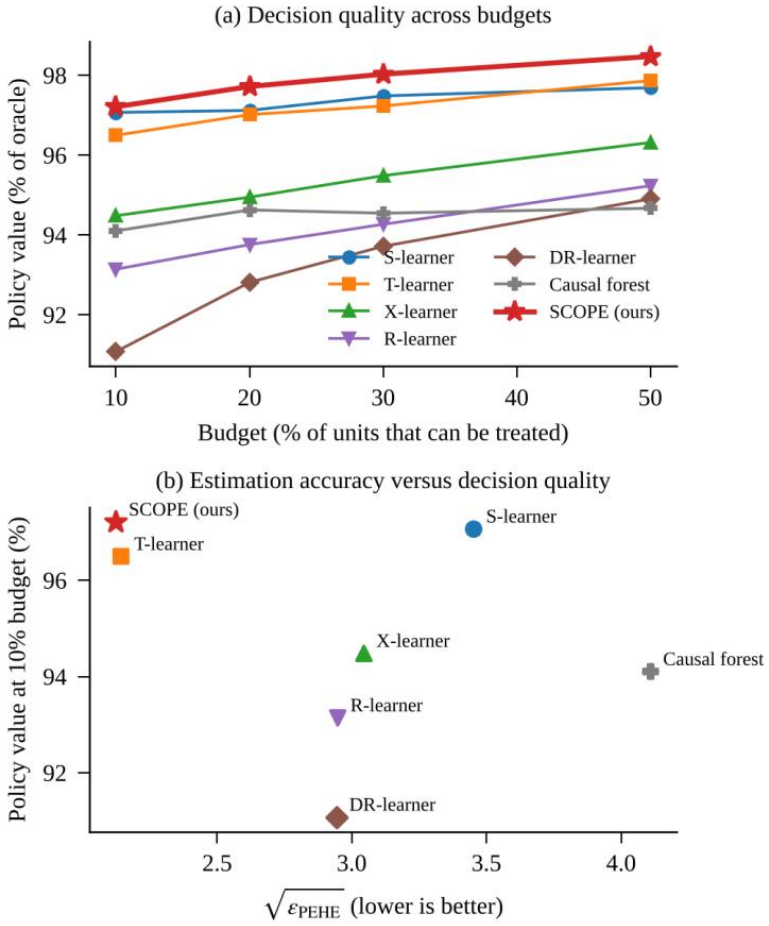



**Figure 2** Results on IHDP over 50 Replications
(a) Policy value relative to the oracle allocation. (b) Estimation error versus decision quality at a 10% budget.

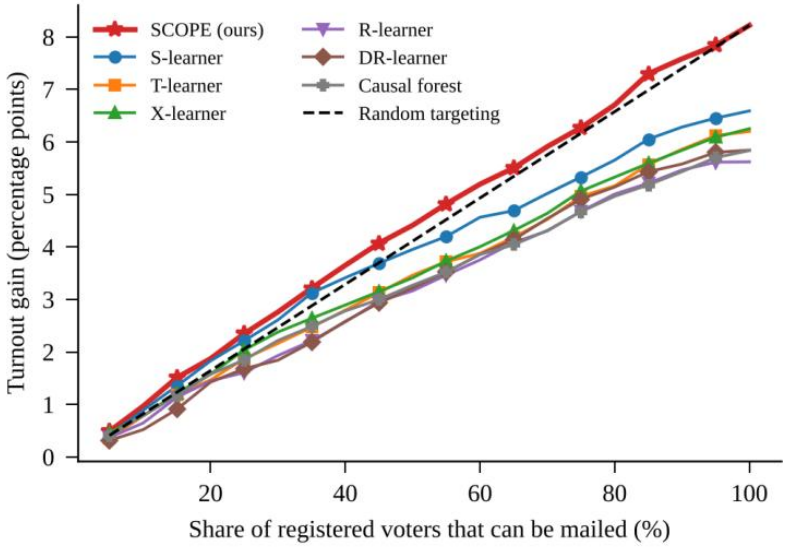



**Figure 3** Budget–Value Frontier on the GOTV Field Experiment
The dashed line sends the most effective message (neighbors) to randomly selected voters.

Table 3 and Figure 3 report the per-capita turnout gain on the 214,107 held-out voters. SCOPE obtains the largest gain at every budget: with a 10% budget it yields 0.968 pp, 8.9% more than the best single learner (S-learner, 0.889 pp) and 17.8% more than random targeting with the most effective message (0.822 pp); with a 50% budget the gain is 4.412 pp versus 3.953 pp for the best single learner. The average effect among the voters selected by SCOPE at 10% is therefore 9.68 pp, compared with 8.22 pp for the neighbors message on average.

Most single learners fall below random targeting. Their noisy estimates for the two weak mailings, whose average effects are small, divert mailings away from the neighbors message. SCOPE's arm-wise stacking resolves this failure mode as predicted by Proposition 2: for the civic-duty arm all learner weights are zero and the intercept equals 1.85 pp, the doubly robust average effect, so the arm is treated as homogeneous; for the Hawthorne arm the stack retains a damped X-learner (weight 0.14); and for the neighbors arm it keeps a shrunken S-learner (weight 0.20) on top of an intercept of 5.78 pp. The budget–value frontier in Figure 3 lies above random targeting at every budget below 100% and shows how the marginal value of additional mailings declines as the budget grows, which informs the choice of the budget itself.

### 5.5 Results on the Multi-Treatment Simulator

**Table 4** Results on the Synthetic Benchmark (3 Seeds)

| Method | √PEHE | V@10% | V@20% | V@30% | V@50% |
|---|---|---|---|---|---|
| S-learner | 0.290±0.008 | 91.8±0.1 | 92.5±0.2 | 93.2±0.3 | 94.1±0.4 |
| T-learner | 0.485±0.011 | 77.9±0.4 | 80.2±0.7 | 81.8±0.4 | 84.4±0.4 |
| X-learner | 0.350±0.010 | 85.2±0.5 | 86.8±0.2 | 88.2±0.3 | 90.1±0.3 |
| R-learner | 0.497±0.016 | 77.2±0.8 | 80.1±1.0 | 81.9±0.9 | 84.4±0.6 |
| DR-learner | 0.584±0.017 | 71.2±3.0 | 75.5±2.2 | 78.2±1.8 | 81.5±1.5 |
| Causal forest | 0.330±0.001 | 85.2±0.7 | 86.6±0.6 | 87.7±0.8 | 89.1±1.0 |
| **SCOPE (ours)** | **0.242±0.014** | **92.4±0.2** | **93.4±0.3** | **94.2±0.4** | **95.2±0.5** |

With three incentives, heterogeneous costs and weak overlap, the gaps between learners widen (Table 4). The DR- and R-learners suffer from the extreme propensity scores, and the causal forest reaches 85.2–89.1% of the oracle value. SCOPE attains 92.41–95.24%, the highest value at every budget, and reduces √PEHE by 16.3% relative to the best single learner (S-learner). The learned weights on the S-learner exceed one for two of the three arms (1.14, 1.13, 0.97), so the stack rescales its shrunken effects, while the negative intercepts correct their level; both adjustments act on the net scores that drive cross-arm decisions under heterogeneous costs.

### 5.6 Ablation Study

**Table 5** Ablation of the Stacking Stage (Policy Value)

| Variant | IHDP 10% | IHDP 30% | Syn. 10% | Syn. 30% | GOTV 10% | GOTV 30% |
|---|---|---|---|---|---|---|
| SCOPE (arm-wise, intercept) | 97.21 | 98.03 | **92.41** | **94.17** | **0.968** | **2.760** |
| Arm-wise, no intercept | 97.47 | **98.19** | 92.24 | 93.98 | 0.967 | 2.758 |
| Shared weights | 97.47 | **98.19** | 92.17 | 93.72 | 0.810 | 2.536 |
| Convex weights | **97.53** | 98.18 | 91.89 | 93.60 | 0.810 | 2.536 |
| Uniform average | 96.42 | 97.42 | 84.82 | 88.18 | 0.892 | 2.298 |
| DR-loss selection | 97.47 | 98.18 | 92.32 | 93.55 | 0.810 | 2.536 |
| Library S only | 96.90 | 97.32 | 92.32 | 93.55 | 0.810 | 2.536 |
| Library T only | 96.91 | 97.73 | 80.17 | 84.02 | 0.881 | 2.439 |
| Library X only | 94.28 | 95.39 | 85.71 | 88.87 | 0.886 | 2.496 |
| Library R only | 95.65 | 96.46 | 79.98 | 83.87 | 0.879 | 2.417 |
| Library DR only | 91.46 | 93.79 | 73.56 | 80.06 | 0.596 | 1.840 |

Table 5 isolates the stacking design and Figure 4 shows the learned weights; GOTV values are in pp and the other values in percent of the oracle. Uniform averaging of the library is clearly inferior on the two multi-treatment problems, which shows that the weights must be learned. Shared or convex weights and the selection of a single learner by the DR loss are competitive on IHDP, where one arm and uniform costs make the allocation invariant to scale and level (Proposition 2), but they reach only 0.810 pp on GOTV, where the arm-wise design is 19.5% higher. The intercept adds 0.17 points at the 10% budget on the simulator, whose arms have heterogeneous costs. The full design is the best variant on both multi-treatment problems and stays within 0.32 points of the best variant on IHDP. The library rows confirm that no single member is uniformly reliable: the S-learner leads on the simulator, whereas the T-, X- and R-learners lead on GOTV.

For the allocation layer we compare three solvers on the stacked scores of the simulator test sets (20,000 units, three incentives). The dual-price bisection reaches an objective within 0.010% of the LP relaxation solved by HiGHS while being about 83 times faster (0.016 s versus 1.50 s), and its policy value matches that of the rounded LP solution. The greedy benefit-to-cost heuristic, which ignores the option of upgrading a unit to a costlier arm, loses up to 0.79 points of policy value at the 50% budget.

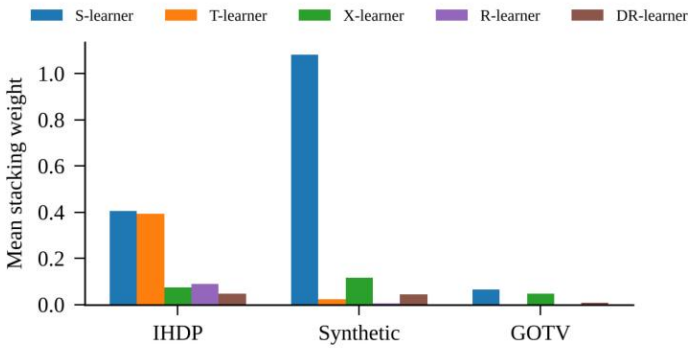



**Figure 4** Average Stacking Weights per Learner
Weights above one indicate that the stack rescales a shrunken learner.

### 5.7 Scalability

**Table 6** Solve Time (s) and Gap to the LP Relaxation (%) of the Allocators

| Units | Dual | Greedy | LP | Dual gap | Greedy gap |
|---|---|---|---|---|---|
| 10,000 | 0.009 | 0.001 | 0.40 | 0.0113 | 0.210 |
| 30,000 | 0.030 | 0.002 | 3.12 | 0.0030 | 0.129 |
| 100,000 | 0.101 | 0.007 | 39.31 | 0.0024 | 0.123 |
| 300,000 | 0.284 | 0.022 | – | – | – |
| 1,000,000 | 1.331 | 0.093 | – | – | – |

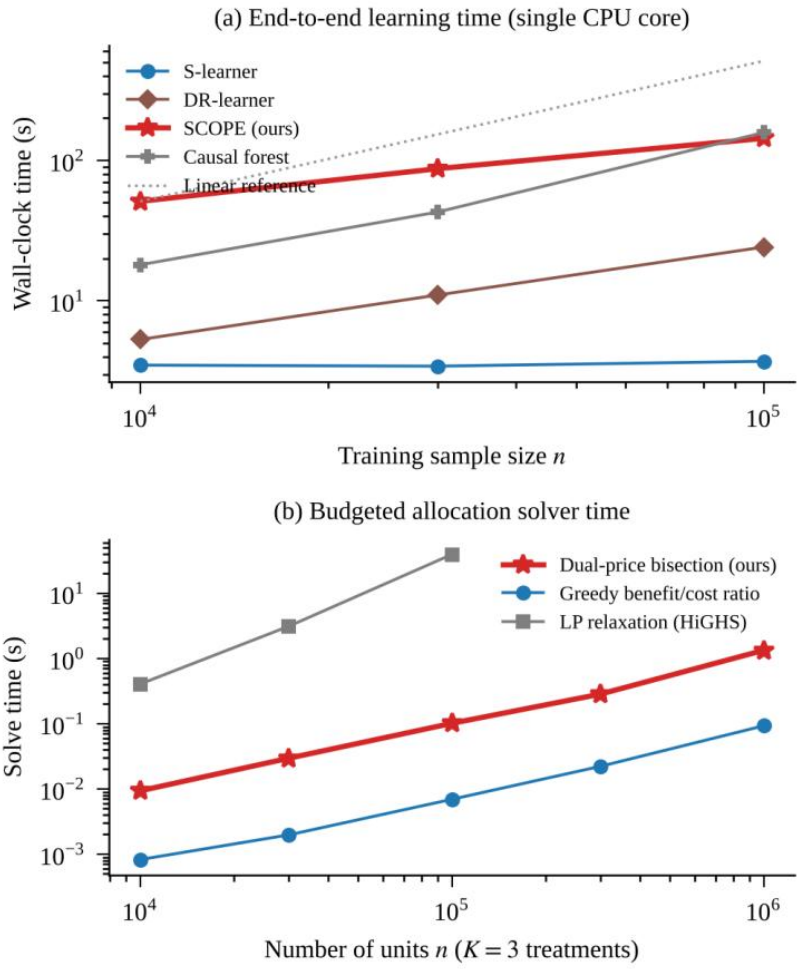



**Figure 5** Scalability on a Single CPU Core
(a) End-to-end training and scoring time versus the number of training units. (b) Solve time of the budgeted allocation with $K = 3$ treatments.

Figure 5 and Table 6 summarize the scalability study on the simulator, using a fixed set of 50,000 deployment units. SCOPE trains on 100,000 units in 144 s on one core, of which 14% is spent on the nuisance cache, 60% on the library

and less than 1% on stacking. Its runtime grows with an empirical log–log slope of 0.45 over the range studied, compared with 0.66 for the DR-learner and 0.94 for the causal forest. At 100,000 units SCOPE is already faster than the causal forest (144 s versus 158 s) while reaching 97.6% versus 92.3% of the oracle value at a 20% budget. Because SCOPE trains a whole library with cross-fitting, its runtime is a constant multiple of that of a single learner (6.0 times the DR-learner at 100,000 units); this factor does not grow with $n$, and the underlying fits are independent jobs that parallelize across cores. The decision quality of SCOPE improves with data, from 91.7% of the oracle value at 10,000 units to 97.6% at 100,000 units for a 20% budget.

The allocation layer scales to 1,000,000 units with three treatments in 1.33 s. The LP relaxation requires 39.3 s at 100,000 units, 389 times longer than the dual method, whereas the dual objective stays within 0.0113% of the LP optimum on every instance in Table 6.

## 6 CONCLUSION

This paper presented SCOPE, a scalable causal machine learning framework that turns heterogeneous treatment effect estimation into budgeted multi-treatment decisions. Its four stages, namely a cross-fitted nuisance cache, a leakage-free meta-learner library, orthogonal arm-wise stacking and dual-price allocation, run in linear time and are supported by an analysis that connects stacking risk, arm-wise calibration and decision regret. Across a semi-synthetic benchmark, a 305,866-voter field experiment and a confounded multi-treatment simulator, SCOPE delivers the highest policy value among all compared methods, raising the turnout gain of a 10% mailing budget by 8.9% over the best single learner, and its allocator solves million-unit instances in seconds with a negligible gap to the linear relaxation. Future work will extend the framework to sequential budget allocation, to multiple resource constraints and to distributed training of the library.

## COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

## REFERENCES

[1] Künzel S R, Sekhon J S, Bickel P J, et al. Metalearners for estimating heterogeneous treatment effects using machine learning. Proceedings of the National Academy of Sciences of the United States of America, 2019, 116(10): 4156-4165.
[2] Nie X, Wager S. Quasi-oracle estimation of heterogeneous treatment effects. Biometrika, 2021, 108(2): 299-319.
[3] Kennedy E H. Towards optimal doubly robust estimation of heterogeneous causal effects. Electronic Journal of Statistics, 2023, 17(2): 3008-3049.
[4] Chernozhukov V, Chetverikov D, Demirer M, et al. Double/debiased machine learning for treatment and structural parameters. The Econometrics Journal, 2018, 21(1): C1-C68.
[5] Foster D J, Syrgkanis V. Orthogonal statistical learning. The Annals of Statistics, 2023, 51(3): 879-908.
[6] Wager S, Athey S. Estimation and inference of heterogeneous treatment effects using random forests. Journal of the American Statistical Association, 2018, 113(523): 1228-1242.
[7] Athey S, Tibshirani J, Wager S. Generalized random forests. The Annals of Statistics, 2019, 47(2): 1148-1178.
[8] Shalit U, Johansson F D, Sontag D. Estimating individual treatment effect: Generalization bounds and algorithms. In: Proceedings of the 34th International Conference on Machine Learning (ICML), PMLR 70, 2017: 3076-3085.
[9] Shi C, Blei D M, Veitch V. Adapting neural networks for the estimation of treatment effects. In: Advances in Neural Information Processing Systems 32 (NeurIPS), 2019.
[10] Curth A, van der Schaar M. In search of insights, not magic bullets: Towards demystification of the model selection dilemma in heterogeneous treatment effect estimation. In: Proceedings of the 40th International Conference on Machine Learning (ICML), PMLR 202, 2023: 6623-6642.
[11] Fernández-Loría C, Provost F. Causal decision making and causal effect estimation are not the same… and why it matters. INFORMS Journal on Data Science, 2022, 1(1): 4-16.
[12] Sinha P, Zoltners A A. The multiple-choice knapsack problem. Operations Research, 1979, 27(3): 503-515.
[13] Zhao Z, Harinen T. Uplift modeling for multiple treatments with cost optimization. In: Proceedings of the IEEE International Conference on Data Science and Advanced Analytics (DSAA), 2019.
[14] Ai M, Li B, Gong H, et al. LBCF: A large-scale budget-constrained causal forest algorithm. In: Proceedings of the ACM Web Conference (WWW), 2022: 2310-2319.
[15] Zhou H, Li S, Jiang G, et al. Direct heterogeneous causal learning for resource allocation problems in marketing. In: Proceedings of the AAAI Conference on Artificial Intelligence, 2023, 37(4): 5446-5454.
[16] Hill J L. Bayesian nonparametric modeling for causal inference. Journal of Computational and Graphical Statistics, 2011, 20(1): 217-240.
[17] Gutierrez P, Gérardy J Y. Causal inference and uplift modelling: A review of the literature. In: Proceedings of the International Conference on Predictive Applications and APIs, PMLR 67, 2017: 1-13.
[18] Alaa A, Ahmad Z, van der Laan M. Conformal meta-learners for predictive inference of individual treatment effects. In: Advances in Neural Information Processing Systems 36 (NeurIPS), 2023.

[19] Kitagawa T, Tetenov A. Who should be treated? Empirical welfare maximization methods for treatment choice. Econometrica, 2018, 86(2): 591-616.
[20] Dudík M, Langford J, Li L. Doubly robust policy evaluation and learning. In: Proceedings of the 28th International Conference on Machine Learning (ICML), 2011: 1097-1104.
[21] Athey S, Wager S. Policy learning with observational data. Econometrica, 2021, 89(1): 133-161.
[22] Zhou Z, Athey S, Wager S. Offline multi-action policy learning: Generalization and optimization. Operations Research, 2023, 71(1): 148-183.
[23] Bertsimas D, Kallus N. From predictive to prescriptive analytics. Management Science, 2020, 66(3): 1025-1044.
[24] Elmachtoub A N, Grigas P. Smart 'predict, then optimize'. Management Science, 2022, 68(1): 9-26.
[25] Mandi J, Kotary J, Berden S, et al. Decision-focused learning: Foundations, state of the art, benchmark and future opportunities. arXiv:2307.13565, 2023.
[26] Breiman L. Stacked regressions. Machine Learning, 1996, 24(1): 49-64.
[27] van der Laan M J, Polley E C, Hubbard A E. Super learner. Statistical Applications in Genetics and Molecular Biology, 2007, 6(1): 25.
[28] Ke G, Meng Q, Finley T, et al. LightGBM: A highly efficient gradient boosting decision tree. In: Advances in Neural Information Processing Systems 30 (NIPS), 2017: 3146-3154.
[29] Gerber A S, Green D P, Larimer C W. Social pressure and voter turnout: Evidence from a large-scale field experiment. American Political Science Review, 2008, 102(1): 33-48.