CAUSALMIX: Data Mixture as Causal Inference for Language Model Training

Zinan Tang[1,2,†] **Yukun Zhang**[2,†] **Shaomian Zheng**[2] **Zhuoshi Pan**[1] **Qizhi Pei**[3]
Dingnan Jin[2] **Jun Zhou**[2] ***Yujun Wang***[*] ***Biqing Huang***[1,*]

[1]Tsinghua University, [2]Ant Group, [3]Renmin University of China

Abstract

In Large Language Model (LLM) training, data mixing plays a pivotal role in determining model performance. Recent methods optimize mixture weights via proxy models, but they rely on the assumption of static data distributions. As a result, when the underlying data pool shifts, these methods require costly retraining from scratch. This limitation restricts their ability to scale seamlessly from small settings to larger data pools and model sizes. In this paper, we propose CAUSALMIX to address this limitation by casting data mixture optimization as a causal inference problem. We formulate the statistical features of the data pool as covariates and the domain mixture as the treatment. After fitting a causal model on 512 runs of `Qwen2.5-0.5B` to estimate the Conditional Average Treatment Effect (CATE), we extrapolate the optimal mixture for an 800K data pool and apply it to train a 7B model. Furthermore, we successfully generalize the framework to long chain-of-thought data on `Qwen3-4B-Base`. By leveraging causal modeling to isolate confounding biases, CAUSALMIX dynamically infers state-dependent optimal data mixtures. Extensive experiments show that the mixture guided by CAUSALMIX consistently improves performance across multiple downstream tasks, outperforming RegMix and other baselines. In addition, we use the CATE Interpreter to provide visual analysis of the learned mixing strategy. Overall, CAUSALMIX offers a causal and interpretable framework for optimizing LLM data mixtures.

1 Introduction

The remarkable capabilities of Large Language Models (LLMs) are driven by the quality and composition of their training data (Zhang et al., 2025a; Kandpal & Raffel, 2025; Tang et al., 2025; Gao et al., 2025). During Supervised Fine-Tuning (SFT), where models are aligned with human intent and specialized for complex tasks, the data mixture, namely the relative proportion of different domains such as instruction following, mathematical reasoning, and coding, has a substantial impact on downstream performance (Li et al., 2025a). However, determining the optimal mixture remains a notoriously challenging problem. One reason is that training LLMs is expensive, making exhaustive grid search over the continuous simplex of mixture weights intractable for large-scale models.

Existing automated data mixing strategies typically approach this problem through the lens of representation learning or proxy modeling. Methods such as RegMix (Liu et al., 2025) optimize data weights by minimizing validation loss on a reference dataset, treating historical training runs as independent samples to fit a global mapping from mixture weights to loss. While effective for pre-training, these loss-centric approaches often falter during SFT (Xu et al., 2026; Li & Kim, 2026; Zhang et al., 2025b). Moreover, global mappings fail to account for the profound impact of the *data state*, namely the inherent complexity, quality, and difficulty of the specific data pool being used. In other words, a single static optimal mixture does not exist (Wang et al., 2025; Tao et al., 2026).

To bridge this gap, we propose CAUSALMIX, a framework that formulates data mixture optimization not

[*]Corresponding Authors.
[†]Equal Contribution.
[‡]Work during research internship at Ant Group.

as a black-box hyperparameter search, but as a *causal marginal return estimation problem*. Instead of seeking a universal mapping from mixture proportions T to absolute performance Y, we treat historical proxy training runs as treatments. By conditioning on the data state X, characterized by metrics such as normalized loss (Shum et al., 2025), entropy (Li et al., 2026), and writing style (Wettig et al., 2024), we ask a localized causal question: *How does a relative change in domain proportions causally affect downstream performance under the current data state?*

Drawing upon Double Machine Learning (DML) (Chernozhukov et al., 2018) and causal forests (Wager & Athey, 2018; Oprescu et al., 2019), CAUSALMIX orthogonalizes the treatment and outcome variables with respect to the data state. This ensures that the estimated marginal returns are isolated from the confounding effects of the data pool's inherent quality. Once the causal direction is identified, we employ a conservative policy update, constrained by a trust region, to adjust the mixture weights.

The causal perspective of CAUSALMIX not only provides a principled optimization objective but also unlocks interpretability and transferability. By analyzing the Conditional Treatment Effects (CATE), we empirically unearth the "skill conflicts" between factual knowledge and complex logical reasoning (Wu et al., 2025; Balappanawar et al., 2025), and demonstrate how data quality thresholds dictate the effectiveness of math and coding data. Furthermore, because CAUSALMIX learns the underlying causal dynamics rather than memorizing a specific dataset, it successfully extrapolates to entirely unseen data pools and larger model architectures without requiring new proxy experiments. Taken together, these results position CAUSALMIX as a principled and practical framework for scalable, interpretable, and transferable data mixture optimization in LLM training.

2 Related works

Data mixture optimization. Data mixture plays an important role in LLM training and strongly affects downstream task performance. Most existing offline methods (Xie et al., 2023b; Albalak et al., 2023; Liu et al., 2025; Fan et al., 2024; Ye et al., 2025; Chen et al., 2025) focus on the pre-training stage, deriving domain weights through proxy models or modeling training loss as a function of the data mixture. In contrast, data mixture optimization for SFT remains relatively underexplored. Existing SFT-oriented methods, such as DMO (Li et al., 2025b) and IDEAL (Ming et al., 2026), still fundamentally use validation loss as the optimization objective. SMART (Renduchintala et al., 2024) is a relatively rare exception that does not directly optimize validation loss; instead, it formulates data selection as two consecutive cardinality-constrained submodular maximization problems.

Causal inference in machine learning. Integrating causal inference with machine learning helps mitigate spurious correlations and distribution shifts in traditional data-driven models (Peters et al., 2017; Schölkopf et al., 2021). This line of research is grounded in the potential outcomes framework (Rubin, 2005; Imbens & Rubin, 2015) and causal graphical models (Pearl, 2009; Spirtes et al., 2000). Recent work has mainly progressed along three directions: improving treatment effect estimation through deep representation learning for confounder control (Shalit et al., 2017; Louizos et al., 2017; Shi et al., 2019) and DML frameworks (Chernozhukov et al., 2018); uncovering latent data-generating structure through differentiable causal discovery (Zheng et al., 2018) and mechanism disentanglement (Bengio et al., 2020); and improving generalization by incorporating causal invariance into objectives (Rojas-Carulla et al., 2018; Arjovsky et al., 2019; Liu et al., 2021).

3 Methodology

We formulate data mixture optimization as a state-conditioned causal marginal return estimation problem. An overview of the CAUSALMIX pipeline is shown in Figure 1.

3.1 Target estimand and identification

Given K data domains and a fixed training budget, a mixture is represented as

$$T = (T_1, \ldots, T_K), \qquad T_k \geq 0, \qquad \sum_{k=1}^{K} T_k = 1.$$

Each training run with a prescribed mixture can be viewed as a data-mixture treatment, and the resulting downstream performance is the corresponding outcome. To simultaneously capture the diminishing marginal returns dictated by empirical scaling laws (Kaplan et al., 2020; Xu et al., 2026) and accommodate the standard geometric transformations for compositional data on a probability simplex (Aitchison, 1982), we define the

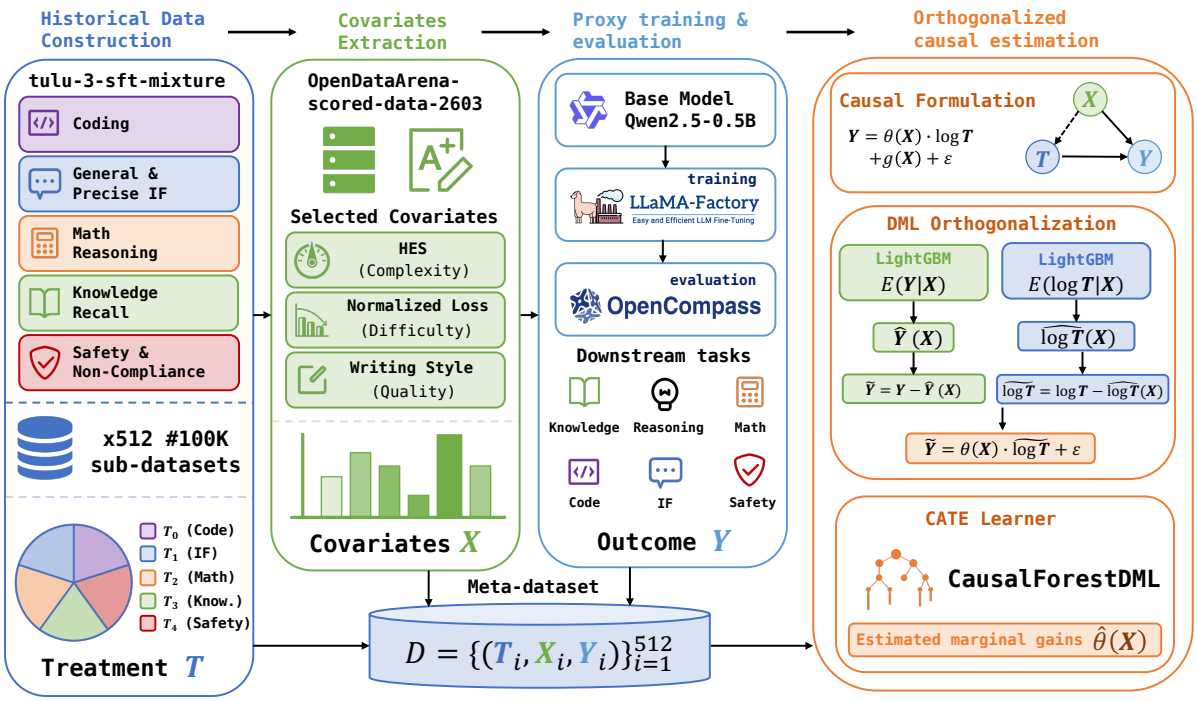

Figure 1: Overview of the CAUSALMIX pipeline. Historical proxy runs provide data-state covariates, mixture assignments, and downstream outcomes, which are used to estimate state-conditioned marginal data returns through orthogonal causal learning.

continuous treatment using a log-mixture representation:

$$Z = \log(T + \varepsilon),$$

where the logarithm is applied element-wise and $\varepsilon > 0$ is a small smoothing constant.

For the i-th historical proxy run, we observe a triplet (X_i, T_i, Y_i), where covariates X_i denotes the data state available before training and evaluation, treatment T_i is the mixture fixed before training, and outcome Y_i is the downstream performance after training. The state X_i is a policy context, such as quality, difficulty, complexity, or stylistic statistics of the data pool; it must not include post-training model information or downstream evaluation results. Let $Y_i(t)$ be the potential outcome that would be obtained if run i were trained with raw mixture t under the same training budget, recipe, sampling rule, and evaluation protocol. We define the conditional response function with respect to the corresponding log-mixture $z = \log(t + \varepsilon)$ as

$$\mu(x, z) = \mathbb{E}[Y(t) \mid X = x].$$

Learning the full response surface $\mu(x, z)$ is difficult because the mixture space is continuous and the number of proxy runs is limited. We therefore focus on the local marginal response within the treatment support covered by historical mixtures. For a given x, we use the partially linear approximation (Chernozhukov et al., 2018; Robinson, 1988; Nie & Wager, 2021)

$$\mu(x, Z) \approx g(x) + \theta_0(x)^\top Z,$$

where $g(x)$ captures the state-dependent baseline performance, and $\theta_0(x) \in \mathbb{R}^K$ is the *state-conditioned marginal data return*. The quantity $\theta_0(x)$ can be understood as a generalized CATE for multidimensional continuous treatments. If $\theta_{0,k}(x) > 0$, increasing the relative proportion of domain k tends to improve downstream performance under state x; if $\theta_{0,k}(x) < 0$, increasing that domain may induce negative transfer. Unlike feature importance in standard supervised learning, $\theta_0(x)$ describes the local causal response of potential outcomes to mixture treatments.

To identify this quantity from proxy runs, we assume consistency and ignorability by design:

$$Y_i = Y_i(T_i), \qquad Y(t) \perp T \mid X.$$

The second condition requires that the mixture-generation mechanism be specified before training and evaluation, and that it does not depend on downstream outcomes, training feedback, or other unobserved information that would systematically affect potential outcomes. Since Z is a deterministic transformation of T, this directly implies $Y(z) \perp Z \mid X$. We also assume local overlap and local smoothness within the historical treatment support. Under these assumptions,

$$\mathbb{E}[Y \mid X = x, Z = z] = \mathbb{E}[Y(z) \mid X = x] = \mu(x, z),$$

so the local marginal return $\theta_0(x)$ is identifiable from proxy mixture experiments. If mixtures are selected adaptively using training or evaluation results, this interpretation should be weakened to a causally motivated marginal-response estimate.

3.2 Orthogonal estimation of marginal returns

Direct regression is not aligned with our objective, because it mixes state-dependent baseline effects with the causal effect of mixture changes. We address this issue using Double Machine Learning (DML) (Chernozhukov et al., 2018), which residualizes both the outcome and the treatment with respect to the covariates to isolate the local causal response. We therefore define the nuisance functions

$$m_0(X) = \mathbb{E}[Y \mid X], \qquad e_0(X) = \mathbb{E}[Z \mid X],$$

and construct residuals

$$\widetilde{Y} = Y - m_0(X), \qquad \widetilde{Z} = Z - e_0(X).$$

The marginal return is estimated from the residualized relation

$$\widetilde{Y} \approx \theta_0(X)^\top \widetilde{Z}.$$

This asks whether deviations in log-mixture proportions beyond their state-conditioned expectation explain performance deviations beyond the state-conditioned baseline.

In practice, the nuisance functions are estimated with cross-fitting (Oprescu et al., 2019; Wager & Athey, 2018): historical proxy runs are split into folds, and each residual is generated by first-stage models trained without the corresponding sample. We then learn a heterogeneous effect model by minimizing the orthogonal loss

$$\hat{\theta} = \arg\min_{\theta} \sum_i \left(\widetilde{Y}_i - \theta(X_i)^\top \widetilde{Z}_i \right)^2.$$

This is an R-loss-style objective: it does not optimize absolute-score prediction, but instead estimates how residual treatment variation explains residual outcome variation.

3.3 From marginal returns to mixture policies

After estimating the target-state marginal return $\hat{\theta}(X_{\text{tar}})$ with respect to log-mixture proportions, we convert it into a feasible raw mixture on the simplex. The guiding principle is simple: domains with larger positive log-mixture marginal returns should receive larger weights, while domains with low or negative marginal returns should not be encouraged to increase.

A deterministic analytical extraction maps positive log-mixture marginal returns to the simplex:

$$T_k^{\text{A}} = \frac{[\hat{\theta}_k(X_{\text{tar}})]_+}{\sum_{j=1}^{K} [\hat{\theta}_j(X_{\text{tar}})]_+}, \qquad [a]_+ = \max(a, 0).$$

The detailed mathematical proof is provided in Appendix B. And a search-based extraction instead evaluates a set of raw candidate mixtures $T^{(1)}, \ldots, T^{(M)}$. Each candidate is transformed into its log-treatment representation $Z^{(m)} = \log(T^{(m)} + \varepsilon)$ before being passed to the fitted causal model. Let $\widehat{S}(Z^{(m)}; X_{\text{tar}})$ denote the predicted score or predicted gain of candidate $T^{(m)}$ at the target state. The final strategy is obtained by averaging the top candidates in the original mixture space:

$$T^{\text{S}} = \frac{1}{K_{\text{top}}} \sum_{m \in \text{Top}} T^{(m)}.$$

This can be viewed as local bagging over high-scoring candidates: instead of relying on a single potentially overestimated mixture, it averages several strong raw-mixture candidates to reduce inference noise, smooth the resulting policy, and enhance generalization.

4 Experiments

In this section, we first describe the experimental setup, then compare CAUSALMIX with strong baselines across different data scales and model sizes, further conduct extension experiments on LongCoT data, and finally present ablation studies. We provide experimental details including introductions of datasets, models, benchmarks and baselines, training and evaluation hyperparameters, and computing costs in Appendix A.

4.1 Experimental setup

Data preparation. We use the `tulu-3-sft-mixture` (Lambert et al., 2025) dataset and adopt the domain partitioning strategy introduced in Tulu 3. Specifically, we consider five domains: Coding, Instruction Following (IF, combining General and Precise IF), Math Reasoning, Knowledge Recall, and Safety & Non-Compliance. We sample 512 sub-datasets, each containing 100K instances, and denote the domain mixture proportions of each sub-dataset as the treatment T. To efficiently extract data features, we leverage `OpenDataArena-scored-data-2603` (OpenDataArena, 2025b; Cai et al., 2025; OpenDataArena, 2025a), which provides pre-computed scores on 30 metrics spanning multiple dimensions such as Diversity, Complexity and Quality. A carefully selected subset of these metrics, namely `Normalized_Loss` (Shum et al., 2025), `Writing_Style` (Wettig et al., 2024), and `HES` (Li et al., 2026), serves as our covariates X. Detailed analysis of this selection is provided in Section 5.3.

Proxy model training and evaluation. We select Qwen2.5-0.5B (Qwen et al., 2025; Yang et al., 2024) as the proxy model and conduct training using `LlamaFactory` (Zheng et al., 2024). For evaluation, we use `OpenCompass` (Contributors, 2023) to assess the models on a diverse suite of downstream tasks aligned with the training domains, following the Tulu 3 evaluation protocol (Lambert et al., 2025). We group the downstream tasks into six capabilities: Knowledge, Reasoning, Math, Coding, IF and Safety. We further partition these benchmarks into Development set $\mathcal{S}_{\text{Dev}}$ and Unseen set $\mathcal{S}_{\text{Uns}}$. We adopt the domain-level micro-average score on $\mathcal{S}_{\text{Dev}}$ as the final outcome Y.

Causal model fitting and inference. We use the `EconML` (Battocchi et al., 2019) framework for causal model fitting and inference. Specifically, we adopt `LightGBM` (Ke et al., 2017) as the first-stage predictor and `CausalForestDML` (Wager & Athey, 2018; Chernozhukov et al., 2018; Oprescu et al., 2019) as the core causal estimator; detailed rationales for these choices are provided in Section 5.1 and Section 5.2. After fitting the causal model on a meta-dataset of 512 historical (X, T, Y) triplets, we set the covariate X to the comprehensive feature profile of the full `tulu-3-sft-mixture` training dataset. We consider two variants: CAUSALMIX-A (Analytical), which directly computes the exact closed-form solution. And CAUSALMIX-S (Search), following the practice of RegMix (Liu et al., 2025), we draw 100,000 candidate mixtures from a Dirichlet distribution and perform inference on these candidates. We then average the top-100 performing mixtures to obtain the final strategy.

Baselines. We compare CAUSALMIX against several representative baselines. These include Grid, which denotes the best mixture proportion empirically identified from the 512 proxy-model runs, as well as existing automated mixing methods including RegMix (Liu et al., 2025), DoReMi (Xie et al., 2023a), ODM (Albalak et al., 2023) and DMO (Li et al., 2025a). To ensure a fair comparison, rather than directly adopting the static mixture proportions reported in the original papers, we re-implement these automated methods and train them on our own historical runs following their official protocols. For DMO, we instead use the mixing ratios reported in its paper.

4.2 Main result

As illustrated in Table 1, CAUSALMIX achieves strong performance on $\textbf{Avg}_{\text{Dev}}$ and also demonstrates strong generalization on $\mathcal{S}_{\text{Uns}}$. Notably, CAUSALMIX-S performs better than CAUSALMIX-A on $\textbf{Avg}_{\text{Uns}}$, which may result from averaging the top-100 candidate mixtures: this procedure can smooth out idiosyncratic variance in individual solutions and thus yield a more robust strategy. To reduce the possibility that the observed gains are due to chance, we conduct repeated comparisons across multiple training data scales, ranging from 100K to 800K. Across these settings, our method always outperforms several baselines, especially the recent SFT-oriented state-of-the-art (SOTA) method DMO. Inspired by the *rank invariance hypothesis* proposed by RegMix (Liu et al., 2025), we further scale the model size to 7B under the 800K data setting and observe a similar performance trend. This cross-scale consistency further supports the effectiveness and robustness of our approach.

Table 1: Performance comparison of different data mixture methods across different model sizes and data scales. The highest average scores are highlighted in bold, and the second-highest are underlined.

Method	Knowledge	Reasoning	Math	Coding	IF	Safety	Avg_{Dev}	Avg_{Uns}
Qwen2.5-0.5B-Instruct	29.90	32.14	35.84	34.10	30.31	11.14	28.90	28.60
Qwen2.5-7B-Instruct	60.24	52.36	54.70	68.67	44.36	60.47	56.80	46.94
Llama-3.1-Tulu-3-8B-SFT	51.63	67.37	56.14	57.91	68.39	46.39	57.97	41.46
Qwen2.5-0.5B, # 100K, tulu-3-sft-mixture								
Equal	27.65	29.80	23.32	33.00	33.83	14.57	27.03	24.43
Grid	28.46	31.98	30.41	33.50	20.15	23.38	27.98	23.45
RegMix	27.29	30.04	25.53	31.28	17.01	32.56	27.28	**26.24**
DoReMi	29.81	31.70	22.90	33.40	35.30	25.34	<u>29.74</u>	24.12
ODM	27.50	30.53	25.41	33.10	26.06	23.62	27.70	23.59
DMO	28.86	30.91	22.51	30.68	35.86	26.19	29.17	<u>26.08</u>
CausalMix-A	28.37	31.31	24.67	29.57	38.63	26.93	**29.91**	23.42
CausalMix-S	27.27	30.23	27.77	34.94	34.94	14.81	27.85	25.90
Qwen2.5-0.5B, # 400K, tulu-3-sft-mixture								
Equal	28.39	32.98	23.96	32.10	37.89	27.78	30.51	24.71
Grid	29.34	31.84	26.68	35.02	16.64	17.26	26.13	27.08
RegMix	28.57	30.19	27.02	34.32	24.21	16.65	26.82	<u>26.07</u>
DoReMi	28.58	31.12	22.14	33.61	39.74	26.56	30.29	25.69
ODM	28.65	30.74	24.35	34.62	30.50	12.61	26.91	25.51
DMO	27.74	31.75	24.25	31.80	38.63	41.62	<u>32.63</u>	25.84
CausalMix-A	28.73	30.63	24.67	30.48	42.51	43.45	**33.41**	24.26
CausalMix-S	29.45	31.30	26.37	32.40	38.82	25.95	30.71	**26.93**
Qwen2.5-0.5B, # 800K, tulu-3-sft-mixture								
Equal	28.07	29.28	21.59	36.06	46.95	25.09	31.78	<u>25.64</u>
Grid	23.94	24.40	29.67	23.81	17.93	22.64	31.17	22.53
RegMix	24.90	29.94	30.25	25.46	25.32	22.52	26.40	22.82
DoReMi	27.70	29.66	20.58	34.53	42.88	32.93	31.38	24.88
ODM	28.59	30.50	25.18	36.06	33.64	10.28	27.37	25.36
DMO	28.07	31.88	22.42	26.90	41.59	41.37	32.04	**26.49**
CausalMix-A	27.93	29.68	23.56	27.76	43.81	50.92	**33.94**	25.02
CausalMix-S	28.31	30.96	27.64	30.97	42.51	36.47	<u>32.81</u>	25.04
Qwen2.5-7B, # 800K, tulu-3-sft-mixture								
Equal	60.85	64.55	59.03	53.61	68.58	53.49	60.02	**49.55**
Grid	59.08	68.77	65.99	61.55	44.92	56.43	59.46	46.37
RegMix	59.60	68.12	63.37	55.76	58.04	55.94	60.14	48.12
DoReMi	58.02	63.20	57.35	57.28	68.21	52.75	59.47	46.19
ODM	60.25	65.69	59.42	44.22	63.59	54.83	58.00	48.00
DMO	59.15	63.70	60.62	54.05	70.24	54.35	60.35	48.98
CausalMix-A	57.14	64.03	58.51	65.52	68.21	57.65	<u>61.84</u>	<u>49.09</u>
CausalMix-S	59.35	62.88	58.63	64.43	67.47	60.95	**62.28**	47.98

4.3 Extension experiments

To rigorously evaluate the transferability of CausalMix, we conduct an extended generalization experiment across disparate data pools and model architectures. Specifically, we repurpose the historical data from `tulu-3-sft-mixture` (Lambert et al., 2025), retain the same covariate selection for X, and define the outcome Y as the average downstream performance in the coding and math domains. Subsequently, we apply the trained causal predictor to the entirely unseen dataset `AM-Thinking-v1-Distilled-math&code` (Tian et al., 2025) to infer the optimal mixture proportions. To validate the effectiveness of these extrapolated weights, we train and evaluate `Qwen3-4B` (Team, 2025a), a model series distinct from the proxy model (the `Qwen2.5` series (Yang et al., 2024)). Empirical evaluations demonstrate that CausalMix consistently achieves the best performance. This robust transferability demonstrates that our CausalMix successfully captures the

Table 2: Performance comparison of different data mixture methods on LongCoT data. The highest scores are highlighted in bold, and the second-highest are underlined.

Method	GSM8K	MATH	Avg_{Math}	HumanEval	MBPP	Avg_{Code}	Avg
Qwen3-4B, # 20K, AM-Thinking-v1-Distilled-Code&Math							
Equal	90.45	56.78	<u>73.62</u>	59.76	48.20	53.98	<u>63.80</u>
Grid	87.34	61.20	74.27	62.80	47.60	55.20	64.74
RegMix	89.61	40.80	65.21	61.59	53.60	57.60	61.40
DoReMi	88.55	42.22	65.39	63.41	53.80	**58.61**	62.00
ODM	88.32	41.16	64.74	63.41	42.20	52.81	58.77
DMO	89.61	54.38	72.00	54.88	55.00	54.94	63.47
CAUSALMIX	88.86	60.58	**74.72**	62.20	55.00	<u>58.60</u>	**66.66**

intrinsic laws of data mixing, enabling seamless extrapolation across datasets and models without costly proxy-model retraining, and further validating its effectiveness on LongCoT data.

4.4 Ablation study

Table 3: Ablation study of the key components in CAUSALMIX. Removing the DML orthogonalization step (*w/o Orth.*) or discarding covariates (*w/o X*) both lead to performance degradation. The highest average scores are highlighted in bold.

Method	Knowledge	Reasoning	Math	Coding	IF	Safety	Avg
Qwen2.5-0.5B, # 800K, tulu-3-sft-mixture							
w/o X	29.27	31.39	29.97	33.41	39.37	36.35	33.29
w/o Orth.	27.41	31.29	24.74	31.90	41.04	37.82	32.66
CAUSALMIX-A	27.93	29.68	23.56	27.76	43.81	50.92	**33.94**
CAUSALMIX-S	28.31	30.96	27.64	30.97	42.51	36.47	32.81
Qwen2.5-7B, # 800K, tulu-3-sft-mixture							
w/o X	60.45	63.95	61.16	55.62	69.69	56.92	61.30
w/o Orth.	59.50	64.66	60.45	45.82	68.76	58.75	59.65
CAUSALMIX-A	57.14	64.03	58.51	65.52	68.21	57.65	61.84
CAUSALMIX-S	59.35	62.88	58.63	64.43	67.47	60.95	**62.28**

We compare CAUSALMIX against two degraded variants, both of which use `LightGBM` as the underlying regressor, to validate the necessity of its key components. (1) *w/o X*. We entirely remove the state covariates, yielding a RegMix-like variant (Liu et al., 2025). Unlike RegMix, however, its optimization target is not validation loss but the average performance on downstream tasks. In this setting, the model reduces to learning a global mapping from treatment to outcome, $\hat{Y} = g(T)$. By attempting to learn this static mapping without conditioning on the data state, this context-agnostic variant becomes highly vulnerable to distribution shifts, leading to the performance degradation. (2) *w/o Orth.* We bypass the DML orthogonalization step and directly concatenate covariates X and treatment T to predict the absolute outcome $\hat{Y}$, i.e., $\hat{Y} = f(X, T)$. As shown in Table 3, this direct regression leads to clear performance degradation, even performing worse than directly fitting T, which further highlights the regularization bias inherent in standard supervised learning.

5 Analysis

In this section, we analyze the choices of the causal estimator and covariates. We further use the CATE model interpreter to provide interpretable insights into the dynamics of data mixing.

5.1 Causal model selection

To identify the most suitable causal estimator, we perform model selection using the R-Scorer (R-loss) metric. The R-Scorer provides a principled and approximately unbiased criterion based onRobinson (1988)'s orthogonalization technique. It enables us to compare different causal estimators by evaluating how well their predicted causal effects $\hat{\theta}(X)$ explain variation in the residual outcomes $\tilde{Y}$ given the residual treatments $\tilde{T}$.

We evaluate a range of causal estimators in the EconML framework that support multidimensional continuous treatments, and report the results in Table 4 (a). Among them, CausalForestDML achieves the best performance. We attribute this advantage to its non-parametric, tree-based recursive partitioning architecture. Unlike linear causal models that impose rigid parametric assumptions, CausalForestDML is suited to capturing the complex interactions between the multidimensional covariates and treatment space. It also naturally accommodates feature saturation and localized heterogeneous effects, making it particularly suitable for the intricate dynamics of data mixing (Wager & Athey, 2018; Chernozhukov et al., 2018; Oprescu et al., 2019).

Table 4: Model selection results for the causal estimator and first-stage predictors. Left (a): performance of candidate causal estimators measured by RScore. Right (b): representative first-stage predictor combinations ranked by RScore. The selected models are highlighted by color.

Model	**RScore** (↑)
LinearDML	+0.1445
SparseLinearDML	-1.7065
CausalForestDML	**+0.1683**
CausalForestDML_Deep	-0.1238
CausalForestDML_Shallow	+0.0207
DML_Poly2_Lasso	+0.1404
DML_Poly2_Ridge	+0.0340
DML_Poly3_Lasso	+0.1533
DML_Poly3_Ridge	+0.0653

Rank	**Y Predictor**	**T Predictor**	**RScore** (↑)	**Time** (s)
1	**LightGBM**	**LightGBM**	**0.1683**	12.9
5	RandomForest	LightGBM	-0.0556	10.3
6	RidgeCV	LightGBM	-0.1408	6.8
7	ElasticNetCV	LightGBM	-0.1681	**5.5**
8	GradientBoosting	LightGBM	-0.1686	15.4
19	RandomForest	RandomForest	-0.2840	18.7
23	GradientBoosting	RandomForest	-0.3241	22.5
24	LassoCV	LightGBM	-0.4098	11.1
25	LightGBM	RandomForest	-0.4323	18.0
29	GradientBoosting	GradientBoosting	-0.4816	10.0

5.2 First-stage predictor selection

The first-stage predictors estimate the conditional expectations $\hat{Y}(X)$ and $\hat{T}(X)$. We evaluate a diverse set of regression algorithms and their combinations for the outcome and treatment models. As summarized in Table 4 (b), with the full results provided in the Appendix, using LightGBM for both models achieves the highest RScore. This configuration substantially outperforms all other standalone regressors as well as linear models. Notably, LightGBM also emerges as the best treatment predictor across all top-ranked configurations. We attribute this strong performance to LightGBM's efficient framework, which handles the multidimensional statistical features while capturing variable interactions without severe overfitting (Ke et al., 2017). Although its computational cost is not the lowest among the candidates, this trade-off is acceptable given the substantial performance gains.

5.3 Covariates selection

In causal inference, covariate selection is of central importance. To efficiently identify the most informative covariates from OpenDataArena-scored-data-2603 (OpenDataArena, 2025a), we randomly sample 64 instances from our 512 historical records as a validation set. Keeping all other hyperparameters fixed, we train distinct causal models with different covariate combinations. We then generate predictions and evaluate them by computing the Spearman rank correlation with the ground-truth scores. We experiment with the vast majority of combinations across different sizes.

As shown in Figure 2, the best performance is achieved with a combination of three covariates. Specifically, HES sums the entropy of the top 0.5% highest-entropy tokens in reasoning traces produced by Qwen3-8B to capture critical decision points and genuine reasoning complexity (Li et al., 2026). Normalized_Loss computes the normalized cross-entropy using Qwen3-8B (Shum et al., 2025), reflecting data predictability and training utility. Finally, Writing_Style evaluates the clarity, coherence, and stylistic quality of the text using QuRater-1.3B (Wettig et al., 2024).

These three metrics naturally correspond to the broader dimensions of data Complexity, Difficulty, and Quality, respectively. This leads to an important finding: effective causal modeling requires controlling for a diverse feature profile rather than focusing on only one aspect of the data. However, incorporating too many covariates degrades performance. We attribute this decline primarily to the limited size of our historical meta-dataset, which makes the causal estimator more vulnerable to the curse of dimensionality. For a fair comparison with prior methods such as RegMix (Liu et al., 2025), we fix the number of proxy models to 512. We expect that scaling up the number of proxy models to support more covariates could further improve performance.

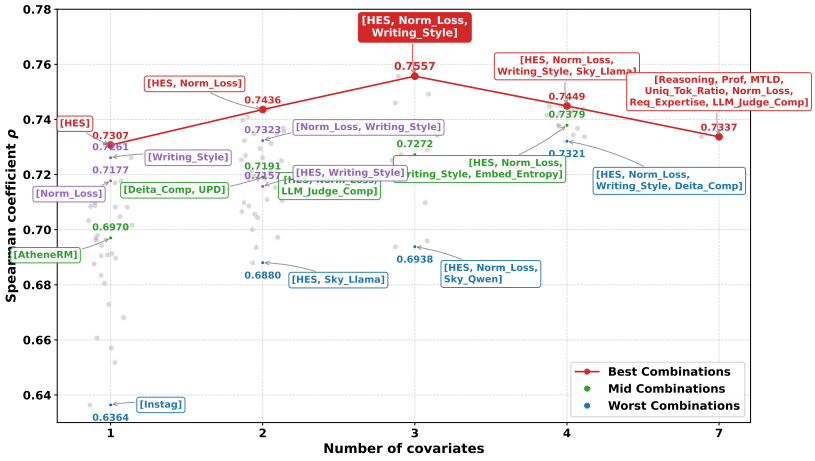

Figure 2: Spearman rank correlation under different covariate combinations.

5.4 CATE Interpreter

We conduct a Tree Interpreter analysis of the trained causal model, as shown in Figure 3. The results show that IF data is the primary driver of downstream alignment, yielding stable positive returns across feature subspaces. In contrast, Knowledge data has negative effects on difficult target data characterized by high `Normalized_Loss` and high `HES`, corroborating the existence of "skill conflicts" between logical reasoning and factual knowledge injection (Wu et al., 2025; Balappanawar et al., 2025). Moreover, the marginal returns of different domains depend strongly on the characteristics of the target data. In low-quality regions, characterized by low `Writing_Style` and low `HES`, complex domains such as Math, Coding, and Safety introduce distributional noise and degrade performance. However, when `Writing_Style` and `HES` is moderate, these domains produce strong synergistic gains, effectively mitigating the performance penalty typically associated with Safety data.

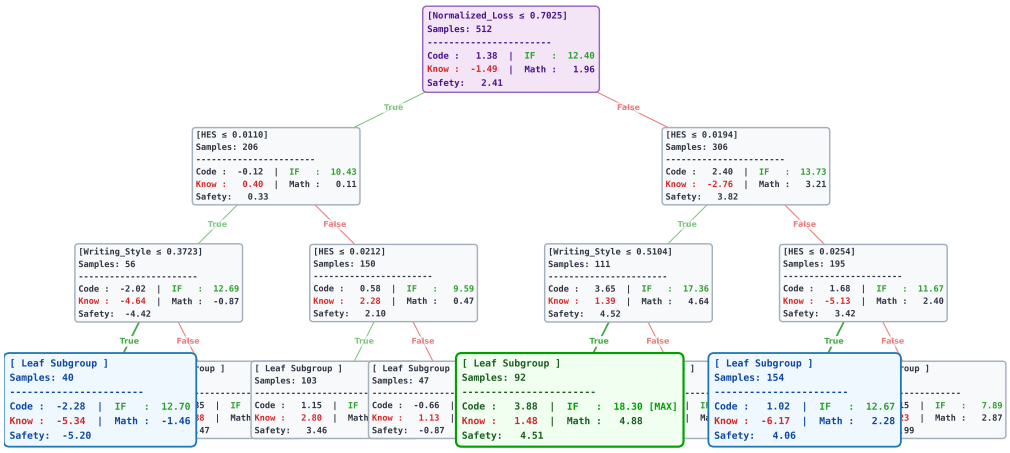

Figure 3: Simplified visualization of the CATE model tree interpreter.

6 Conclusion

In this work, we introduced CAUSALMIX, a framework that shifts SFT data mixture optimization from static validation-loss minimization to state-conditioned causal marginal return estimation. By treating historical proxy training runs as causal treatments and combining orthogonalized estimation with a conservative trust-region policy, CAUSALMIX isolates the marginal utility of domain proportions from the confounding effects of the underlying data state. Extensive experiments show that our approach consistently outperforms strong baselines across different model scales and data budgets, while also exhibiting strong transferability

to unseen LongCoT data pools. Furthermore, the interpretable insights derived from our causal framework, including quantified skill conflicts between factual knowledge injection and complex logical reasoning, provide a principled foundation for future research on understanding and optimizing the dynamics of LLM training.

References

John Aitchison. The statistical analysis of compositional data. *Journal of the Royal Statistical Society: Series B (Methodological)*, 44(2):139–160, 1982.

Alon Albalak, Liangming Pan, Colin Raffel, and William Yang Wang. Efficient online data mixing for language model pre-training, 2023. URL https://arxiv.org/abs/2312.02406.

Martin Arjovsky, Léon Bottou, Ishaan Gulrajani, and David Lopez-Paz. Invariant risk minimization. *arXiv preprint arXiv:1907.02893*, 2019.

Jacob Austin, Augustus Odena, Maxwell Nye, Maarten Bosma, Henryk Michalewski, David Dohan, Ellen Jiang, Carrie Cai, Michael Terry, Quoc Le, and Charles Sutton. Program synthesis with large language models, 2021. URL https://arxiv.org/abs/2108.07732.

Ishwar B Balappanawar, Vamshi Krishna Bonagiri, Anish R Joishy, Manas Gaur, Krishnaprasad Thirunarayan, and Ponnurangam Kumaraguru. If pigs could fly... can llms logically reason through counterfactuals? *arXiv preprint arXiv:2505.22318*, 2025.

Keith Battocchi, Eleanor Dillon, Maggie Hei, Greg Lewis, Paul Oka, Miruna Oprescu, and Vasilis Syrgkanis. EconML: A Python Package for ML-Based Heterogeneous Treatment Effects Estimation. https://github.com/py-why/EconML, 2019. Version 0.x.

Edward Beeching, Shengyi Costa Huang, Albert Jiang, Jia Li, Benjamin Lipkin, Zihan Qina, Kashif Rasul, Ziju Shen, Roman Soletskyi, and Lewis Tunstall. Numinamath 7b tir. https://huggingface.co/AI-MO/NuminaMath-7B-TIR, 2024.

Yoshua Bengio, Tristan Deleu, Nasim Rahaman, Nan Rosemary Ke, Sebastien Lachapelle, Olexa Bilaniuk, Anirudh Goyal, and Christopher Pal. A meta-transfer objective for learning to disentangle causal mechanisms. In *International Conference on Learning Representations*, 2020. URL https://openreview.net/forum?id=ryxWIgBFPS.

Faeze Brahman, Sachin Kumar, Vidhisha Balachandran, Pradeep Dasigi, Valentina Pyatkin, Abhilasha Ravichander, Sarah Wiegreffe, Nouha Dziri, Khyathi Chandu, Jack Hessel, Yulia Tsvetkov, Noah A. Smith, Yejin Choi, and Hannaneh Hajishirzi. The art of saying no: Contextual noncompliance in language models. In *The Thirty-eight Conference on Neural Information Processing Systems Datasets and Benchmarks Track*, 2024. URL https://openreview.net/forum?id=f1UL4wNlw6.

Mengzhang Cai, Xin Gao, Yu Li, Honglin Lin, Zheng Liu, Zhuoshi Pan, Qizhi Pei, Xiaoran Shang, Mengyuan Sun, Zinan Tang, et al. Opendataarena: A fair and open arena for benchmarking post-training dataset value. *arXiv preprint arXiv:2512.14051*, 2025.

Mark Chen, Jerry Tworek, Heewoo Jun, Qiming Yuan, Henrique Ponde de Oliveira Pinto, Jared Kaplan, Harri Edwards, Yuri Burda, Nicholas Joseph, Greg Brockman, Alex Ray, Raul Puri, Gretchen Krueger, Michael Petrov, Heidy Khlaaf, Girish Sastry, Pamela Mishkin, Brooke Chan, Scott Gray, Nick Ryder, Mikhail Pavlov, Alethea Power, Lukasz Kaiser, Mohammad Bavarian, Clemens Winter, Philippe Tillet, Felipe Petroski Such, Dave Cummings, Matthias Plappert, Fotios Chantzis, Elizabeth Barnes, Ariel Herbert-Voss, William Hebgen Guss, Alex Nichol, Alex Paino, Nikolas Tezak, Jie Tang, Igor Babuschkin, Suchir Balaji, Shantanu Jain, William Saunders, Christopher Hesse, Andrew N. Carr, Jan Leike, Josh Achiam, Vedant Misra, Evan Morikawa, Alec Radford, Matthew Knight, Miles Brundage, Mira Murati, Katie Mayer, Peter Welinder, Bob McGrew, Dario Amodei, Sam McCandlish, Ilya Sutskever, and Wojciech Zaremba. Evaluating large language models trained on code, 2021. URL https://arxiv.org/abs/2107.03374.

Mayee F Chen, Michael Y. Hu, Nicholas Lourie, Kyunghyun Cho, and Christopher Re. Aioli: A unified optimization framework for language model data mixing. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=sZGZJhaNSe.

Victor Chernozhukov, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins. Double/debiased machine learning for treatment and structural parameters, 2018.

Karl Cobbe, Vineet Kosaraju, Mohammad Bavarian, Mark Chen, Heewoo Jun, Lukasz Kaiser, Matthias Plappert, Jerry Tworek, Jacob Hilton, Reiichiro Nakano, Christopher Hesse, and John Schulman. Training verifiers to solve math word problems, 2021. URL https://arxiv.org/abs/2110.14168.

OpenCompass Contributors. Opencompass: A universal evaluation platform for foundation models. https://github.com/open-compass/opencompass, 2023.

Yuntian Deng, Wenting Zhao, Jack Hessel, Xiang Ren, Claire Cardie, and Yejin Choi. WildVis: Open source visualizer for million-scale chat logs in the wild. In Delia Irazu Hernandez Farias, Tom Hope, and Manling Li (eds.), *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pp. 497–506, Miami, Florida, USA, November 2024. Association for Computational Linguistics. doi: 10.18653/v1/2024.emnlp-demo.50. URL https://aclanthology.org/2024.emnlp-demo.50/.

Dheeru Dua, Yizhong Wang, Pradeep Dasigi, Gabriel Stanovsky, Sameer Singh, and Matt Gardner. DROP: A reading comprehension benchmark requiring discrete reasoning over paragraphs. In Jill Burstein, Christy Doran, and Thamar Solorio (eds.), *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pp. 2368–2378, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics. doi: 10.18653/v1/N19-1246. URL https://aclanthology.org/N19-1246/.

Simin Fan, Matteo Pagliardini, and Martin Jaggi. Doge: Domain reweighting with generalization estimation. In *International Conference on Machine Learning*, pp. 12895–12915. PMLR, 2024.

Xin Gao, Qizhi Pei, Zinan Tang, Yu Li, Honglin Lin, Jiang Wu, Lijun Wu, and Conghui He. A strategic coordination framework of small LMs matches large LMs in data synthesis. In Wanxiang Che, Joyce Nabende, Ekaterina Shutova, and Mohammad Taher Pilehvar (eds.), *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 11552–11570, Vienna, Austria, July 2025. Association for Computational Linguistics. ISBN 979-8-89176-251-0. doi: 10.18653/v1/2025.acl-long.566. URL https://aclanthology.org/2025.acl-long.566/.

Seungju Han, Kavel Rao, Allyson Ettinger, Liwei Jiang, Bill Yuchen Lin, Nathan Lambert, Yejin Choi, and Nouha Dziri. Wildguard: Open one-stop moderation tools for safety risks, jailbreaks, and refusals of LLMs. In *The Thirty-eight Conference on Neural Information Processing Systems Datasets and Benchmarks Track*, 2024. URL https://openreview.net/forum?id=Ich4tv4202.

Thomas Hartvigsen, Saadia Gabriel, Hamid Palangi, Maarten Sap, Dipankar Ray, and Ece Kamar. ToxiGen: A large-scale machine-generated dataset for adversarial and implicit hate speech detection. In Smaranda Muresan, Preslav Nakov, and Aline Villavicencio (eds.), *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 3309–3326, Dublin, Ireland, May 2022. Association for Computational Linguistics. doi: 10.18653/v1/2022.acl-long.234. URL https://aclanthology.org/2022.acl-long.234/.

Chaoqun He, Renjie Luo, Yuzhuo Bai, Shengding Hu, Zhen Thai, Junhao Shen, Jinyi Hu, Xu Han, Yujie Huang, Yuxiang Zhang, Jie Liu, Lei Qi, Zhiyuan Liu, and Maosong Sun. OlympiadBench: A challenging benchmark for promoting AGI with olympiad-level bilingual multimodal scientific problems. In Lun-Wei Ku, Andre Martins, and Vivek Srikumar (eds.), *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 3828–3850, Bangkok, Thailand, August 2024. Association for Computational Linguistics. doi: 10.18653/v1/2024.acl-long.211. URL https://aclanthology.org/2024.acl-long.211/.

Dan Hendrycks, Collin Burns, Steven Basart, Andy Zou, Mantas Mazeika, Dawn Song, and Jacob Steinhardt. Measuring massive multitask language understanding. In *International Conference on Learning Representations*, 2021a. URL https://openreview.net/forum?id=d7KBjmI3GmQ.

Dan Hendrycks, Collin Burns, Saurav Kadavath, Akul Arora, Steven Basart, Eric Tang, Dawn Song, and Jacob Steinhardt. Measuring mathematical problem solving with the MATH dataset. In *Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 2)*, 2021b. URL https://openreview.net/forum?id=7Bywt2mQsCe.

Guido W Imbens and Donald B Rubin. *Causal inference in statistics, social, and biomedical sciences*. Cambridge university press, 2015.

Yunjie Ji, Xiaoyu Tian, Sitong Zhao, Haotian Wang, Shuaiting Chen, Yiping Peng, Han Zhao, and Xiangang Li. Am-thinking-v1: Advancing the frontier of reasoning at 32b scale, 2025. URL https://arxiv.org/abs/2505.08311.

Liwei Jiang, Kavel Rao, Seungju Han, Allyson Ettinger, Faeze Brahman, Sachin Kumar, Niloofar Mireshghallah, Ximing Lu, Maarten Sap, Yejin Choi, and Nouha Dziri. Wildteaming at scale: From in-the-wild jailbreaks to (adversarially) safer language models. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024. URL https://openreview.net/forum?id=n5R6TvBVcX.

Nikhil Kandpal and Colin Raffel. Position: The most expensive part of an LLM *should* be its training data. In *Forty-second International Conference on Machine Learning Position Paper Track*, 2025. URL https://openreview.net/forum?id=L6RpQ1h4Nx.

Jared Kaplan, Sam McCandlish, Tom Henighan, Tom B Brown, Benjamin Chess, Rewon Child, Scott Gray, Alec Radford, Jeffrey Wu, and Dario Amodei. Scaling laws for neural language models. *arXiv preprint arXiv:2001.08361*, 2020.

Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. Lightgbm: a highly efficient gradient boosting decision tree. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, pp. 3149–3157, Red Hook, NY, USA, 2017. Curran Associates Inc. ISBN 9781510860964.

Andreas Köpf, Yannic Kilcher, Dimitri von Rütte, Sotiris Anagnostidis, Zhi Rui Tam, Keith Stevens, Abdullah Barhoum, Duc Minh Nguyen, Oliver Stanley, Richárd Nagyfi, Shahul ES, Sameer Suri, David Alexandrovich Glushkov, Arnav Varma Dantuluri, Andrew Maguire, Christoph Schuhmann, Huu Nguyen, and Alexander Julian Mattick. Openassistant conversations - democratizing large language model alignment. In *Thirty-seventh Conference on Neural Information Processing Systems Datasets and Benchmarks Track*, 2023. URL https://openreview.net/forum?id=VSJotgbPHF.

Nathan Lambert, Jacob Morrison, Valentina Pyatkin, Shengyi Huang, Hamish Ivison, Faeze Brahman, Lester James Validad Miranda, Alisa Liu, Nouha Dziri, Xinxi Lyu, Yuling Gu, Saumya Malik, Victoria Graf, Jena D. Hwang, Jiangjiang Yang, Ronan Le Bras, Oyvind Tafjord, Christopher Wilhelm, Luca Soldaini, Noah A. Smith, Yizhong Wang, Pradeep Dasigi, and Hannaneh Hajishirzi. Tulu 3: Pushing frontiers in open language model post-training. In *Second Conference on Language Modeling*, 2025. URL https://openreview.net/forum?id=i1uGbfHHpH.

Jianwei Li and Jung-Eun Kim. Superficial safety alignment hypothesis. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=9yS40pO1RF.

Peng Li, Yeye He, Dror Yashar, Weiwei Cui, Song Ge, Haidong Zhang, Danielle Rifinski Fainman, Dongmei Zhang, and Surajit Chaudhuri. Table-gpt: Table fine-tuned gpt for diverse table tasks. *Proc. ACM Manag. Data*, 2(3), May 2024. doi: 10.1145/3654979. URL https://doi.org/10.1145/3654979.

Xiaoyuan Li, Yubo Ma, Chengpeng Li, Keqin Bao, Yiyao Yu, Wenjie Wang, Fuli Feng, and Dayiheng Liu. Unified data selection for LLM reasoning, 2026. URL https://openreview.net/forum?id=heVn5cNfje.

Yuan Li, Zhengzhong Liu, and Eric Xing. Data mixing optimization for supervised fine-tuning of large language models. In *Proceedings of the 42nd International Conference on Machine Learning*, ICML'25. JMLR.org, 2025a.

Yuan Li, Zhengzhong Liu, and Eric P. Xing. Data mixing optimization for supervised fine-tuning of large language models. In *Forty-second International Conference on Machine Learning*, 2025b. URL https://openreview.net/forum?id=19kqoNoc2N.

Stephanie Lin, Jacob Hilton, and Owain Evans. TruthfulQA: Measuring how models mimic human falsehoods. In Smaranda Muresan, Preslav Nakov, and Aline Villavicencio (eds.), *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 3214–3252, Dublin, Ireland, May 2022. Association for Computational Linguistics. doi: 10.18653/v1/2022.acl-long.229. URL https://aclanthology.org/2022.acl-long.229/.

Jiashuo Liu, Zheyan Shen, Yue He, Xingxuan Zhang, Renzhe Xu, Han Yu, and Peng Cui. Towards out-of-distribution generalization: A survey. *arXiv preprint arXiv:2108.13624*, 2021.

Jiawei Liu, Chunqiu Steven Xia, Yuyao Wang, and LINGMING ZHANG. Is your code generated by chatGPT really correct? rigorous evaluation of large language models for code generation. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023. URL https://openreview.net/forum?id=1qvx610Cu7.

Qian Liu, Xiaosen Zheng, Niklas Muennighoff, Guangtao Zeng, Longxu Dou, Tianyu Pang, Jing Jiang, and Min Lin. Regmix: Data mixture as regression for language model pre-training. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=5BjQOUXq7i.

Christos Louizos, Uri Shalit, Joris M Mooij, David Sontag, Richard Zemel, and Max Welling. Causal effect inference with deep latent-variable models. *Advances in neural information processing systems*, 30, 2017.

Ziyang Luo, Can Xu, Pu Zhao, Qingfeng Sun, Xiubo Geng, Wenxiang Hu, Chongyang Tao, Jing Ma, Qingwei Lin, and Daxin Jiang. Wizardcoder: Empowering code large language models with evol-instruct. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=UnUwSIgK5W.

Chenlin Ming, Chendi Qu, Qizhi Pei, Zhuoshi Pan, Yu Li, Xiaoming Duan, Lijun Wu, and Conghui He. IDEAL: Data equilibrium adaptation for multi-capability language model alignment. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=n9wS0Hdvri.

Xinkun Nie and Stefan Wager. Quasi-oracle estimation of heterogeneous treatment effects. *Biometrika*, 108(2): 299–319, 2021.

OpenDataArena. Opendataarena-scored-data, 2025a. Hugging Face dataset.

OpenDataArena. OpenDataArena-Tool. https://github.com/OpenDataArena/OpenDataArena-Tool, 2025b. URL https://github.com/OpenDataArena/OpenDataArena-Tool. GitHub repository.

Miruna Oprescu, Vasilis Syrgkanis, and Zhiwei Steven Wu. Orthogonal random forest for causal inference. In *International Conference on Machine Learning*, pp. 4932–4941. PMLR, 2019.

Judea Pearl. *Causality*. Cambridge university press, 2009.

Jonas Peters, Dominik Janzing, and Bernhard Scholkopf. *Elements of causal inference: foundations and learning algorithms*. MIT press, 2017.

Valentina Pyatkin, Saumya Malik, Victoria Graf, Hamish Ivison, Shengyi Huang, Pradeep Dasigi, Nathan Lambert, and Hannaneh Hajishirzi. Generalizing verifiable instruction following. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems Datasets and Benchmarks Track*, 2026. URL https://openreview.net/forum?id=yfYgwjj5F8.

Qwen, :, An Yang, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chengyuan Li, Dayiheng Liu, Fei Huang, Haoran Wei, Huan Lin, Jian Yang, Jianhong Tu, Jianwei Zhang, Jianxin Yang, Jiaxi Yang, Jingren Zhou, Junyang Lin, Kai Dang, Keming Lu, Keqin Bao, Kexin Yang, Le Yu, Mei Li, Mingfeng Xue, Pei Zhang, Qin Zhu, Rui Men, Runji Lin, Tianhao Li, Tianyi Tang, Tingyu Xia, Xingzhang Ren, Xuancheng Ren, Yang Fan, Yang Su, Yichang Zhang, Yu Wan, Yuqiong Liu, Zeyu Cui, Zhenru Zhang, and Zihan Qiu. Qwen2.5 technical report, 2025. URL https://arxiv.org/abs/2412.15115.

Nazneen Rajani, Lewis Tunstall, Edward Beeching, Nathan Lambert, Alexander M. Rush, and Thomas Wolf. No robots. https://huggingface.co/datasets/HuggingFaceH4/no_robots, 2023.

David Rein, Betty Li Hou, Asa Cooper Stickland, Jackson Petty, Richard Yuanzhe Pang, Julien Dirani, Julian Michael, and Samuel R. Bowman. GPQA: A graduate-level google-proof q&a benchmark. In *First Conference on Language Modeling*, 2024. URL https://openreview.net/forum?id=Ti67584b98.

H S V N S Kowndinya Renduchintala, Sumit Bhatia, and Ganesh Ramakrishnan. SMART: Submodular data mixture strategy for instruction tuning. In Lun-Wei Ku, Andre Martins, and Vivek Srikumar (eds.), *Findings of the Association for Computational Linguistics: ACL 2024*, pp. 12916–12934, Bangkok, Thailand, August 2024. Association for Computational Linguistics. doi: 10.18653/v1/2024.findings-acl.766. URL https://aclanthology.org/2024.findings-acl.766/.

Peter M Robinson. Root-n-consistent semiparametric regression. *Econometrica: journal of the Econometric Society*, pp. 931–954, 1988.

Mateo Rojas-Carulla, Bernhard Schölkopf, Richard Turner, and Jonas Peters. Invariant models for causal transfer learning. *Journal of Machine Learning Research*, 19(36):1–34, 2018.

Donald B Rubin. Causal inference using potential outcomes: Design, modeling, decisions. *Journal of the American statistical Association*, 100(469):322–331, 2005.

Bernhard Schölkopf, Francesco Locatello, Stefan Bauer, Nan Rosemary Ke, Nal Kalchbrenner, Anirudh Goyal, and Yoshua Bengio. Toward causal representation learning. *Proceedings of the IEEE*, 109(5):612–634, 2021.

Uri Shalit, Fredrik D Johansson, and David Sontag. Estimating individual treatment effect: generalization bounds and algorithms. In *International conference on machine learning*, pp. 3076–3085. PMLR, 2017.

Claudia Shi, David Blei, and Victor Veitch. Adapting neural networks for the estimation of treatment effects. *Advances in neural information processing systems*, 32, 2019.

Kashun Shum, Yuzhen Huang, Hongjian Zou, Qi Ding, Yixuan Liao, Xiaoxin Chen, Qian Liu, and Junxian He. Predictive data selection: The data that predicts is the data that teaches. *arXiv preprint arXiv:2503.00808*, 2025.

Shivalika Singh, Freddie Vargus, Daniel D'souza, Börje F. Karlsson, Abinaya Mahendiran, Wei-Yin Ko, Herumb Shandilya, Jay Patel, Deividas Mataciunas, Laura O'Mahony, Mike Zhang, Ramith Hettiarachchi, Joseph Wilson, Marina Machado, Luisa Moura, Dominik Krzemiński, Hakimeh Fadaei, Irem Ergun, Ifeoma Okoh, Aisha Alaagib, Oshan Mudannayake, Zaid Alyafeai, Vu Chien, Sebastian Ruder, Surya Guthikonda, Emad Alghamdi, Sebastian Gehrmann, Niklas Muennighoff, Max Bartolo, Julia Kreutzer, Ahmet Üstün, Marzieh Fadaee, and Sara Hooker. Aya dataset: An open-access collection for multilingual instruction tuning. In Lun-Wei Ku, Andre Martins, and Vivek Srikumar (eds.), *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 11521–11567, Bangkok, Thailand, August 2024. Association for Computational Linguistics. doi: 10.18653/v1/2024.acl-long.620. URL https://aclanthology.org/2024.acl-long.620/.

Peter Spirtes, Clark N Glymour, and Richard Scheines. *Causation, prediction, and search*. MIT press, 2000.

Aarohi Srivastava, Abhinav Rastogi, Abhishek Rao, Abu Awal Md Shoeb, Abubakar Abid, Adam Fisch, Adam R. Brown, Adam Santoro, Aditya Gupta, Adrià Garriga-Alonso, Agnieszka Kluska, Aitor Lewkowycz, Akshat Agarwal, Alethea Power, Alex Ray, Alex Warstadt, Alexander W. Kocurek, Ali Safaya, Ali Tazarv, Alice Xiang, Alicia Parrish, Allen Nie, Aman Hussain, Amanda Askell, Amanda Dsouza, Ambrose Slone, Ameet Rahane, Anantharaman S. Iyer, Anders Johan Andreassen, Andrea Madotto, Andrea Santilli, Andreas Stuhlmüller, Andrew M. Dai, Andrew La, Andrew Kyle Lampinen, Andy Zou, Angela Jiang, Angelica Chen, Anh Vuong, Animesh Gupta, Anna Gottardi, Antonio Norelli, Anu Venkatesh, Arash Gholamidavoodi, Arfa Tabassum, Arul Menezes, Arun Kirubarajan, Asher Mullokandov, Ashish Sabharwal, Austin Herrick, Avia Efrat, Aykut Erdem, Ayla Karakaş, B. Ryan Roberts, Bao Sheng Loe, Barret Zoph, Bartłomiej Bojanowski, Batuhan Özyurt, Behnam Hedayatnia, Behnam Neyshabur, Benjamin Inden, Benno Stein, Berk Ekmekci, Bill Yuchen Lin, Blake Howald, Bryan Orinion, Cameron Diao, Cameron Dour, Catherine Stinson, Cedrick Argueta, Cesar Ferri, Chandan Singh, Charles Rathkopf, Chenlin Meng, Chitta Baral, Chiyu Wu, Chris Callison-Burch, Christopher Waites, Christian Voigt, Christopher D Manning, Christopher Potts, Cindy Ramirez, Clara E. Rivera, Clemencia Siro, Colin Raffel, Courtney Ashcraft, Cristina Garbacea, Damien Sileo, Dan Garrette, Dan Hendrycks, Dan Kilman, Dan Roth, C. Daniel Freeman, Daniel Khashabi, Daniel Levy, Daniel Moseguí González, Danielle Perszyk, Danny Hernandez, Danqi Chen, Daphne Ippolito, Dar Gilboa, David Dohan, David Drakard, David Jurgens, Debajyoti Datta, Deep Ganguli, Denis Emelin, Denis Kleyko, Deniz Yuret, Derek Chen, Derek Tam, Dieuwke Hupkes, Diganta Misra, Dilyar Buzan, Dimitri Coelho Mollo, Diyi Yang, Dong-Ho Lee, Dylan Schrader, Ekaterina Shutova, Ekin Dogus Cubuk, Elad Segal, Eleanor Hagerman, Elizabeth Barnes, Elizabeth Donoway, Ellie Pavlick, Emanuele Rodolà, Emma Lam, Eric Chu, Eric Tang, Erkut Erdem, Ernie Chang, Ethan A Chi, Ethan Dyer, Ethan Jerzak, Ethan Kim, Eunice Engefu Manyasi, Evgenii Zheltonozhskii, Fanyue Xia, Fatemeh Siar, Fernando Martínez-Plumed, Francesca Happé, Francois Chollet, Frieda Rong, Gaurav Mishra, Genta Indra Winata, Gerard de Melo, Germàn Kruszewski, Giambattista Parascandolo, Giorgio Mariani, Gloria Xinyue Wang, Gonzalo Jaimovitch-Lopez, Gregor Betz, Guy Gur-Ari, Hana Galijasevic, Hannah Kim, Hannah Rashkin, Hannaneh Hajishirzi, Harsh Mehta, Hayden Bogar, Henry Francis Anthony Shevlin, Hinrich Schuetze, Hiromu Yakura, Hongming Zhang, Hugh Mee Wong, Ian Ng, Isaac Noble, Jaap Jumelet, Jack Geissinger, Jackson Kernion, Jacob Hilton, Jaehoon Lee, Jaime Fernández Fisac, James B Simon, James Koppel, James Zheng, James Zou, Jan Kocon, Jana Thompson, Janelle Wingfield, Jared Kaplan, Jarema Radom, Jascha Sohl-Dickstein, Jason Phang, Jason Wei, Jason Yosinski, Jekaterina Novikova, Jelle Bosscher, Jennifer Marsh, Jeremy Kim, Jeroen Taal, Jesse Engel, Jesujoba Alabi, Jiacheng Xu, Jiaming Song, Jillian Tang, Joan Waweru, John Burden, John Miller, John U. Balis, Jonathan Batchelder, Jonathan Berant, Jörg

Frohberg, Jos Rozen, Jose Hernandez-Orallo, Joseph Boudeman, Joseph Guerr, Joseph Jones, Joshua B. Tenenbaum, Joshua S. Rule, Joyce Chua, Kamil Kanclerz, Karen Livescu, Karl Krauth, Karthik Gopalakrishnan, Katerina Ignatyeva, Katja Markert, Kaustubh Dhole, Kevin Gimpel, Kevin Omondi, Kory Wallace Mathewson, Kristen Chiafullo, Ksenia Shkaruta, Kumar Shridhar, Kyle McDonell, Kyle Richardson, Laria Reynolds, Leo Gao, Li Zhang, Liam Dugan, Lianhui Qin, Lidia Contreras-Ochando, Louis-Philippe Morency, Luca Moschella, Lucas Lam, Lucy Noble, Ludwig Schmidt, Luheng He, Luis Oliveros-Colón, Luke Metz, Lütfi Kerem Senel, Maarten Bosma, Maarten Sap, Maartje Ter Hoeve, Maheen Farooqi, Manaal Faruqui, Mantas Mazeika, Marco Baturan, Marco Marelli, Marco Maru, Maria Jose Ramirez-Quintana, Marie Tolkiehn, Mario Giulianelli, Martha Lewis, Martin Potthast, Matthew L Leavitt, Matthias Hagen, Mátyás Schubert, Medina Orduna Baitemirova, Melody Arnaud, Melvin McElrath, Michael Andrew Yee, Michael Cohen, Michael Gu, Michael Ivanitskiy, Michael Starritt, Michael Strube, Michał Swędrowski, Michele Bevilacqua, Michihiro Yasunaga, Mihir Kale, Mike Cain, Mimee Xu, Mirac Suzgun, Mitch Walker, Mo Tiwari, Mohit Bansal, Moin Aminnaseri, Mor Geva, Mozhdeh Gheini, Mukund Varma T, Nanyun Peng, Nathan Andrew Chi, Nayeon Lee, Neta Gur-Ari Krakover, Nicholas Cameron, Nicholas Roberts, Nick Doiron, Nicole Martinez, Nikita Nangia, Niklas Deckers, Niklas Muennighoff, Nitish Shirish Keskar, Niveditha S. Iyer, Noah Constant, Noah Fiedel, Nuan Wen, Oliver Zhang, Omar Agha, Omar Elbaghdadi, Omer Levy, Owain Evans, Pablo Antonio Moreno Casares, Parth Doshi, Pascale Fung, Paul Pu Liang, Paul Vicol, Pegah Alipoormolabashi, Peiyuan Liao, Percy Liang, Peter W Chang, Peter Eckersley, Phu Mon Htut, Pinyu Hwang, Piotr Miłkowski, Piyush Patil, Pouya Pezeshkpour, Priti Oli, Qiaozhu Mei, Qing Lyu, Qinlang Chen, Rabin Banjade, Rachel Etta Rudolph, Raefer Gabriel, Rahel Habacker, Ramon Risco, Raphaël Millière, Rhythm Garg, Richard Barnes, Rif A. Saurous, Riku Arakawa, Robbe Raymaekers, Robert Frank, Rohan Sikand, Roman Novak, Roman Sitelew, Ronan Le Bras, Rosanne Liu, Rowan Jacobs, Rui Zhang, Russ Salakhutdinov, Ryan Andrew Chi, Seungjae Ryan Lee, Ryan Stovall, Ryan Teehan, Rylan Yang, Sahib Singh, Saif M. Mohammad, Sajant Anand, Sam Dillavou, Sam Shleifer, Sam Wiseman, Samuel Gruetter, Samuel R. Bowman, Samuel Stern Schoenholz, Sanghyun Han, Sanjeev Kwatra, Sarah A. Rous, Sarik Ghazarian, Sayan Ghosh, Sean Casey, Sebastian Bischoff, Sebastian Gehrmann, Sebastian Schuster, Sepideh Sadeghi, Shadi Hamdan, Sharon Zhou, Shashank Srivastava, Sherry Shi, Shikhar Singh, Shima Asaadi, Shixiang Shane Gu, Shubh Pachchigar, Shubham Toshniwal, Shyam Upadhyay, Shyamolima Shammie Debnath, Siamak Shakeri, Simon Thormeyer, Simone Melzi, Siva Reddy, Sneha Priscilla Makini, Soo-Hwan Lee, Spencer Torene, Sriharsha Hatwar, Stanislas Dehaene, Stefan Divic, Stefano Ermon, Stella Biderman, Stephanie Lin, Stephen Prasad, Steven Piantadosi, Stuart Shieber, Summer Misherghi, Svetlana Kiritchenko, Swaroop Mishra, Tal Linzen, Tal Schuster, Tao Li, Tao Yu, Tariq Ali, Tatsunori Hashimoto, Te-Lin Wu, Théo Desbordes, Theodore Rothschild, Thomas Phan, Tianle Wang, Tiberius Nkinyili, Timo Schick, Timofei Kornev, Titus Tunduny, Tobias Gerstenberg, Trenton Chang, Trishala Neeraj, Tushar Khot, Tyler Shultz, Uri Shaham, Vedant Misra, Vera Demberg, Victoria Nyamai, Vikas Raunak, Vinay Venkatesh Ramasesh, vinay uday prabhu, Vishakh Padmakumar, Vivek Srikumar, William Fedus, William Saunders, William Zhang, Wout Vossen, Xiang Ren, Xiaoyu Tong, Xinran Zhao, Xinyi Wu, Xudong Shen, Yadollah Yaghoobzadeh, Yair Lakretz, Yangqiu Song, Yasaman Bahri, Yejin Choi, Yichi Yang, Sophie Hao, Yifu Chen, Yonatan Belinkov, Yu Hou, Yufang Hou, Yuntao Bai, Zachary Seid, Zhuoye Zhao, Zijian Wang, Zijie J. Wang, Zirui Wang, and Ziyi Wu. Beyond the imitation game: Quantifying and extrapolating the capabilities of language models. *Transactions on Machine Learning Research*, 2023. ISSN 2835-8856. URL https://openreview.net/forum?id=uyTL5Bvosj. Featured Certification.

Zinan Tang, Xin Gao, Qizhi Pei, Zhuoshi Pan, Mengzhang Cai, Jiang Wu, Conghui He, and Lijun Wu. Middo: Model-informed dynamic data optimization for enhanced LLM fine-tuning via closed-loop learning. In Christos Christodoulopoulos, Tanmoy Chakraborty, Carolyn Rose, and Violet Peng (eds.), *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pp. 6871–6891, Suzhou, China, November 2025. Association for Computational Linguistics. ISBN 979-8-89176-332-6. doi: 10.18653/v1/2025.emnlp-main.350. URL https://aclanthology.org/2025.emnlp-main.350/.

Chaofan Tao, Zhuo Li, Xiao-Hui Li, Jierun Chen, Weike Jin, Haoli Bai, Lifeng Shang, and Lu Hou. Modalmix: Optimizing multimodal data mixtures with compute-dependent regression, 2026. URL https://openreview.net/forum?id=R1HcIN90A1.

Qwen Team. Qwen3 technical report, 2025a. URL https://arxiv.org/abs/2505.09388.

Qwen Team. Qwq-32b: Embracing the power of reinforcement learning, March 2025b. URL https://qwenlm.github.io/blog/qwq-32b/.

Xiaoyu Tian, Yunjie Ji, Haotian Wang, Shuaiting Chen, Sitong Zhao, Yiping Peng, Han Zhao, and Xiangang

Li. Not all correct answers are equal: Why your distillation source matters, 2025. URL https://arxiv.org/abs/2505.14464.

Shubham Toshniwal, Wei Du, Ivan Moshkov, Branislav Kisacanin, Alexan Ayrapetyan, and Igor Gitman. Openmathinstruct-2: Accelerating AI for math with massive open-source instruction data. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=mTCbq2QssD.

David Wadden, Kejian Shi, Jacob Morrison, Alan Li, Aakanksha Naik, Shruti Singh, Nitzan Barzilay, Kyle Lo, Tom Hope, Luca Soldaini, Shannon Zejiang Shen, Doug Downey, Hannaneh Hajishirzi, and Arman Cohan. SciRIFF: A resource to enhance language model instruction-following over scientific literature. In Christos Christodoulopoulos, Tanmoy Chakraborty, Carolyn Rose, and Violet Peng (eds.), *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pp. 6072–6109, Suzhou, China, November 2025. Association for Computational Linguistics. ISBN 979-8-89176-332-6. doi: 10.18653/v1/2025.emnlp-main.310. URL https://aclanthology.org/2025.emnlp-main.310/.

Stefan Wager and Susan Athey. Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, 113(523):1228–1242, 2018.

Yifan Wang, Binbin Liu, Fengze Liu, Yuanfan Guo, Jiyao Deng, Xuecheng Wu, Weidong Zhou, Xiaohuan Zhou, and Taifeng Wang. Tikmix: Take data influence into dynamic mixture for language model pre-training, 2025. URL https://arxiv.org/abs/2508.17677.

Yubo Wang, Xueguang Ma, Ge Zhang, Yuansheng Ni, Abhranil Chandra, Shiguang Guo, Weiming Ren, Aaran Arulraj, Xuan He, Ziyan Jiang, Tianle Li, Max Ku, Kai Wang, Alex Zhuang, Rongqi Fan, Xiang Yue, and Wenhu Chen. MMLU-pro: A more robust and challenging multi-task language understanding benchmark. In *The Thirty-eight Conference on Neural Information Processing Systems Datasets and Benchmarks Track*, 2024. URL https://openreview.net/forum?id=y10DM6R2r3.

Jason Wei, Maarten Bosma, Vincent Zhao, Kelvin Guu, Adams Wei Yu, Brian Lester, Nan Du, Andrew M. Dai, and Quoc V Le. Finetuned language models are zero-shot learners. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=gEZrGCozdqR.

Alexander Wettig, Aatmik Gupta, Saumya Malik, and Danqi Chen. Qurating: selecting high-quality data for training language models. In *Proceedings of the 41st International Conference on Machine Learning*, ICML'24. JMLR.org, 2024.

Juncheng Wu, Sheng Liu, Haoqin Tu, Hang Yu, Xiaoke Huang, James Zou, Cihang Xie, and Yuyin Zhou. Knowledge or reasoning? a close look at how llms think across domains. *arXiv preprint arXiv:2506.02126*, 2025.

Sang Michael Xie, Hieu Pham, Xuanyi Dong, Nan Du, Hanxiao Liu, Yifeng Lu, Percy Liang, Quoc V. Le, Tengyu Ma, and Adams Wei Yu. Doremi: optimizing data mixtures speeds up language model pretraining. In *Proceedings of the 37th International Conference on Neural Information Processing Systems*, NIPS '23, Red Hook, NY, USA, 2023a. Curran Associates Inc.

Sang Michael Xie, Hieu Pham, Xuanyi Dong, Nan Du, Hanxiao Liu, Yifeng Lu, Percy S Liang, Quoc V Le, Tengyu Ma, and Adams Wei Yu. Doremi: Optimizing data mixtures speeds up language model pretraining. *Advances in Neural Information Processing Systems*, 36:69798–69818, 2023b.

Chengyin Xu, Kaiyuan Chen, Xiao Li, Ke Shen, and Chenggang Li. Unveiling downstream performance scaling of LLMs: A clustering-based perspective. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=3HrDPUi4jx.

An Yang, Baosong Yang, Binyuan Hui, Bo Zheng, Bowen Yu, Chang Zhou, Chengpeng Li, Chengyuan Li, Dayiheng Liu, Fei Huang, Guanting Dong, Haoran Wei, Huan Lin, Jialong Tang, Jialin Wang, Jian Yang, Jianhong Tu, Jianwei Zhang, Jianxin Ma, Jin Xu, Jingren Zhou, Jinze Bai, Jinzheng He, Junyang Lin, Kai Dang, Keming Lu, Keqin Chen, Kexin Yang, Mei Li, Mingfeng Xue, Na Ni, Pei Zhang, Peng Wang, Ru Peng, Rui Men, Ruize Gao, Runji Lin, Shijie Wang, Shuai Bai, Sinan Tan, Tianhang Zhu, Tianhao Li, Tianyu Liu, Wenbin Ge, Xiaodong Deng, Xiaohuan Zhou, Xingzhang Ren, Xinyu Zhang, Xipin Wei, Xuancheng Ren, Yang Fan, Yang Yao, Yichang Zhang, Yu Wan, Yunfei Chu, Yuqiong Liu, Zeyu Cui, Zhenru Zhang, and Zhihao Fan. Qwen2 technical report. *arXiv preprint arXiv:2407.10671*, 2024.

Jiasheng Ye, Peiju Liu, Tianxiang Sun, Jun Zhan, Yunhua Zhou, and Xipeng Qiu. Data mixing laws: Optimizing data mixtures by predicting language modeling performance. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=jjCB27TMK3.

Bolin Zhang, Jiahao Wang, Qianlong Du, Jiajun Zhang, Zhiying Tu, and Dianhui Chu. A survey on data selection for llm instruction tuning. *Journal of Artificial Intelligence Research*, 83, 2025a.

Guanhua Zhang, Ricardo Dominguez-Olmedo, and Moritz Hardt. Train-before-test harmonizes language model rankings. In *NeurIPS 2025 Workshop on Evaluating the Evolving LLM Lifecycle: Benchmarks, Emergent Abilities, and Scaling*, 2025b. URL https://openreview.net/forum?id=GhgsQTb8p9.

Wenting Zhao, Xiang Ren, Jack Hessel, Claire Cardie, Yejin Choi, and Yuntian Deng. Wildchat: 1m chatGPT interaction logs in the wild. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=Bl8u7ZRlbM.

Xun Zheng, Bryon Aragam, Pradeep K Ravikumar, and Eric P Xing. Dags with no tears: Continuous optimization for structure learning. *Advances in neural information processing systems*, 31, 2018.

Yaowei Zheng, Richong Zhang, Junhao Zhang, Yanhan Ye, Zheyan Luo, Zhangchi Feng, and Yongqiang Ma. Llamafactory: Unified efficient fine-tuning of 100+ language models. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 3: System Demonstrations)*, Bangkok, Thailand, 2024. Association for Computational Linguistics. URL http://arxiv.org/abs/2403.13372.

Wanjun Zhong, Ruixiang Cui, Yiduo Guo, Yaobo Liang, Shuai Lu, Yanlin Wang, Amin Saied, Weizhu Chen, and Nan Duan. AGIEval: A human-centric benchmark for evaluating foundation models. In Kevin Duh, Helena Gomez, and Steven Bethard (eds.), *Findings of the Association for Computational Linguistics: NAACL 2024*, pp. 2299–2314, Mexico City, Mexico, June 2024. Association for Computational Linguistics. doi: 10.18653/v1/2024.findings-naacl.149. URL https://aclanthology.org/2024.findings-naacl.149/.

Jeffrey Zhou, Tianjian Lu, Swaroop Mishra, Siddhartha Brahma, Sujoy Basu, Yi Luan, Denny Zhou, and Le Hou. Instruction-following evaluation for large language models. *arXiv preprint arXiv:2311.07911*, 2023.

A Experimental details

A.1 Datasets

We evaluate CAUSALMIX on two SFT datasets.

`tulu-3-sft-mixture` (Lambert et al., 2025) is used to train the Tulu 3 series of models. It contains 939,344 samples spanning seven domains: General (`Tulu 3 Hardcoded`, `OpenAssistant Guanaco` (Köpf et al., 2023), `No Robots` (Rajani et al., 2023), and `WildChat GPT-4` (Zhao et al., 2024; Deng et al., 2024)), Knowledge Recall (`FLAN v2` (Wei et al., 2022), `SciRIFF` (Wadden et al., 2025), and `TableGPT` (Li et al., 2024)), Math Reasoning (`Tulu 3 Persona MATH`, `Tulu 3 Persona GSM`, `Tulu 3 Persona Algebra`, `OpenMathInstruct 2` (Toshniwal et al., 2025), and `NuminaMath-TIR` (Beeching et al., 2024)), Coding (`Tulu 3 Persona Python` and `Evol CodeAlpaca` (Luo et al., 2024)), Safety & Non-Compliance (`CoCoNot` (Brahman et al., 2024), `Tulu 3 WildJailbreak` (Jiang et al., 2024), and `Tulu 3 WildGuardMix` (Han et al., 2024)), Multilingual (`Aya` (Singh et al., 2024)), and Precise IF (`Tulu 3 Persona IF`). We exclude the multilingual subset in our experiments.

`AM-Thinking-v1-Distilled` (Tian et al., 2025) is a reasoning dataset distilled from `AM-Thinking-v1` (Ji et al., 2025). It contains high-quality, automatically verified responses generated from a shared set of 1.89 million queries spanning a wide range of reasoning domains. Its format and verification pipeline allow for direct comparison and seamless integration into downstream tasks. It is intended to support the development of open-source language models with strong reasoning abilities. In our experiments, we use the code and math subsets.

We obtain the data-state covariates from `OpenDataArena-scored-data-2603` (OpenDataArena, 2025b; Cai et al., 2025; OpenDataArena, 2025a).

`OpenDataArena-scored-data-2603` (OpenDataArena, 2025b; Cai et al., 2025; OpenDataArena, 2025a) is a scored SFT dataset collection comprising 63 high-quality instruction-following datasets with nearly 25 million samples. Its core value lies in its 30-dimensional scoring scheme: each sample is evaluated on metrics such as `Normalized_Loss` (Shum et al., 2025), `Writing_Style` (Wettig et al., 2024), `HES` (Li et al., 2026), and 27 others, enabling fine-grained data selection for filtering, curriculum learning, and mixture optimization.

A.2 Models

We use `Qwen2.5-0.5B` as the proxy model, scale the learned mixture strategy up to `Qwen2.5-7B`, and further conduct an extension experiment on `Qwen3-4B-Base`.

`Qwen2.5` (Qwen et al., 2025) is a series of large language models developed by Qwen. It includes both base and instruction-tuned models ranging from 0.5B to 72B parameters. Compared with `Qwen2` (Yang et al., 2024), `Qwen2.5` offers substantially more knowledge and stronger coding and mathematical reasoning capabilities, partly due to specialized expert models in these domains. It also improves instruction following, long-form generation (over 8K tokens), structured-data understanding (e.g., tables), and structured output generation, especially in JSON format. In addition, it is more robust to diverse system prompts, which improves role-play and controllability in chatbot settings. `Qwen2.5` supports contexts of up to 128K tokens and can generate up to 8K tokens, and it supports more than 29 languages, including Chinese, English, French, Spanish, Portuguese, German, Italian, Russian, Japanese, Korean, Vietnamese, Thai, and Arabic. In our experiments, we use the smallest model, `Qwen2.5-0.5B`, as the proxy and scale to the widely used `Qwen2.5-7B`.

`Qwen3` (Team, 2025a) is a newer generation of large language models than `Qwen2.5` in the Qwen series, offering a comprehensive suite of dense and mixture-of-experts (MoE) models. Built on extensive pretraining, `Qwen3` provides substantial advances in reasoning, instruction following, agent capabilities, and multilingual support. Its key features include seamless switching between a thinking mode for complex reasoning, mathematics, and coding and a non-thinking mode for efficient general-purpose dialogue within a single model, enabling strong performance across a wide range of scenarios. It also substantially improves reasoning performance, surpassing previous QwQ (Team, 2025b) models in thinking mode and `Qwen2.5`-Instruct models in non-thinking mode on mathematics, code generation, and commonsense reasoning. In addition, `Qwen3` shows stronger human preference alignment, with better performance in creative writing, role-playing, multi-turn dialogue, and instruction following, resulting in a more natural and engaging conversational experience. It also offers strong agent capabilities, enabling effective integration with external tools in both thinking and non-thinking modes and achieving leading performance among open-source models on complex agentic tasks. Finally, it supports more than 100 languages and dialects and demonstrates strong multilingual instruction-following and translation capabilities. In our extension experiments, we use the 4B dense model.

A.3 Benchmarks

Following Tulu 3, we assess model performance on multiple tasks and corresponding benchmarks, including Knowledge (MMLU (Hendrycks et al., 2021a), MMLU-Pro (Wang et al., 2024), GPQA (Rein et al., 2024)), Reasoning (BBH (Srivastava et al., 2023), DROP (Dua et al., 2019), AGIEval (Zhong et al., 2024)), Math (GSM8K (Cobbe et al., 2021), MATH (Hendrycks et al., 2021b), OlympiadBench (He et al., 2024)), Code (HumanEval (Chen et al., 2021), HumanEval+ (Liu et al., 2023), MBPP (Austin et al., 2021)), Instruction Following (IFEval (Zhou et al., 2023), IFBench (Pyatkin et al., 2026)), and Safety (ToxiGen (Hartvigsen et al., 2022), TruthfulQA (Lin et al., 2022)).

MMLU (Hendrycks et al., 2021a) is heterogeneous with respect to the reasoning skills required to answer the questions, including instances that require basic factual recall as well as those that demand logical reasoning and problem-solving skills. Following Tulu 3, we use a zero-shot chain-of-thought (CoT) setting that asks the model to "summarize" its reasoning before answering the question. We compute the macro average over all subjects in MMLU as the final task metric.

MMLU-Pro (Wang et al., 2024) is a 10-way multiple-choice extension of the MMLU dataset. We use essentially the same prompt and answer extraction procedure as in our AGIEval setup, adjusting only the number of answer choices.

GPQA (Rein et al., 2024) is a set of very challenging multiple-choice questions written by domain experts in biology, physics, and chemistry. We use the same zero-shot prompt and answer extraction procedure as for AGIEval.

BBH (BigBench-Hard) (Srivastava et al., 2023) contains challenging reasoning problems for which models benefit from step-by-step reasoning. We use the default setting of OpenCompass (Contributors, 2023).

DROP (Dua et al., 2019) is a reading comprehension task that requires discrete reasoning. We use the default setting of OpenCompass.

AGIEval (English subset) (Zhong et al., 2024) includes the English-language subset of the AGIEval benchmark, specifically the following multiple-choice tasks: *aqua-rat*, *logiqa-en*, *lsat-ar*, *lsat-lr*, *lsat-rc*, *sat-en*, *sat-math*, and *gaokao-english*. We formulate the task using a simple zero-shot CoT prompt that encourages concise reasoning ending with a clearly stated answer choice. The model's answer choice is extracted by first matching the requested format, with fallback patterns if the format is not followed precisely. Specifically, we first look for the exact phrase indicated in the prompt ("`Therefore, the answer is [ANSWER]`") and take the last such match. If that fails, we look for a sequence of softer variants, such as "`answer is [ANSWER]`" or "`answer: [ANSWER]`", before falling back to the last parenthesized letter found; if that also fails, we use the last stand-alone capital letter.

GSM8K (Cobbe et al., 2021) contains grade-school math word problems. We use the default setting of OpenCompass.

MATH (Hendrycks et al., 2021b) contains problems from mathematics competitions spanning various categories, such as algebra and calculus. We use the default setting of OpenCompass. We compute the macro average across subsections to obtain the final task metric.

OlympiadBench (He et al., 2024) is an Olympiad-level bilingual multimodal scientific benchmark featuring 8,476 problems from Olympiad-level mathematics and physics competitions, including the Chinese college entrance exam. We evaluate only the English math subset and use the same evaluation logic as for MATH.

HumanEval (Chen et al., 2021) and **HumanEval+** (Liu et al., 2023) evaluate models' ability to complete Python code from docstrings. HumanEval+ uses a more rigorous evaluation procedure than the original HumanEval benchmark, with additional tests. We use the default setting of OpenCompass.

MBPP (Austin et al., 2021) contains 974 programming tasks designed to be solvable by entry-level programmers. We use the default setting of OpenCompass.

IFEval (Zhou et al., 2023) evaluates the instruction-following ability of models in a setting where each instruction corresponds to constraints such that it can be programmatically verified whether the outputs satisfy those constraints. We use the default setting of OpenCompass and measure prompt-level accuracy in the loose evaluation setting.

IFBench (Pyatkin et al., 2026) is designed to evaluate generalization in precise instruction following on 58 new, diverse, and challenging verifiable out-of-domain constraints. We use the default setting of OpenCompass.

ToxiGen (Hartvigsen et al., 2022) is a large-scale machine-generated dataset of 274k toxic and benign statements about 13 minority groups. We use a zero-shot setting with unnormalized accuracy.

TruthfulQA (Lin et al., 2022) is a benchmark for measuring whether a language model generates truthful answers to questions. The benchmark comprises 817 questions spanning 38 categories, including health, law, finance, and politics. We use the test split of `mc1` in a zero-shot setting.

We further partition these benchmarks into the development set $\mathcal{S}_{\text{Dev}}$ and the unseen set $\mathcal{S}_{\text{Uns}}$. $\mathcal{S}_{\text{Dev}}$ comprises MMLU, MMLU-Pro, BBH, DROP, GSM8K, MATH, HumanEval, MBPP, IFEval, and TruthfulQA, while $\mathcal{S}_{\text{Uns}}$ consists of GPQA, AGIEval, OlympiadBench, HumanEval+, IFBench, and ToxiGen.

A.4 Baselines

We compare CAUSALMIX with recent methods for offline data mixture optimization.

RegMix (Liu et al., 2025) is designed to automatically identify a high-performing data mixture by formulating the problem as a regression task. It trains many small models on diverse data mixtures, uses regression to predict the performance of unseen mixtures, and applies the best predicted mixture to train a large-scale model with orders of magnitude more compute.

DoReMi (Xie et al., 2023a) first trains a small proxy model using group distributionally robust optimization (Group DRO) over domains to obtain domain weights (mixture proportions) without access to downstream tasks. It then resamples the dataset according to these domain weights and trains a larger full-sized model.

ODM (Albalak et al., 2023) combines elements of both data selection and data mixing. Based on multi-armed bandit algorithms, ODM optimizes the data mixing proportions during training.

DMO (Li et al., 2025a) frames data mixing as an optimization problem and introduces a method designed to minimize validation loss. DMO parameterizes the loss by modeling effective data transfer and leveraging scaling laws for fine-tuning.

A.5 Computing costs

We train 512 proxy models with 0.5B parameters, each on 100K SFT examples. The average sequence length is approximately 4096 tokens, and the total estimated FLOPs are 5.53×10^{20}. Because our method is state-aware, it maintains strong generalization when transferred to out-of-distribution (OOD) data, as demonstrated by the extension experiments in Section 4.3, without requiring the proxy models to be retrained.

For fair comparison, we use the same proxy-model configuration for baseline methods that require proxy training. Specifically, for RegMix (Liu et al., 2025), we also use 512 proxy models. For DoReMi (Xie et al., 2023a), we use a single proxy model. For ODM (Albalak et al., 2023), we use a single model to determine the data mixing proportions and training order. For DMO (Li et al., 2025a), since it is trained on the same data, we directly use the mixture proportions reported in the original paper.

A.6 Hyperparameters

All random seeds in our experiments are set to 42, and all experiments are conducted on NVIDIA H800 GPUs.

Training. For `Qwen2.5-0.5B` and `Qwen2.5-7B`, we follow DMO (Li et al., 2025a); for `Qwen3-4B-Base`, we follow OpenDataArena (Cai et al., 2025). All training hyperparameters are listed in Table 5.

Evaluation. All evaluation hyperparameters are listed in Table 6. For `Qwen2.5-0.5B` and `Qwen2.5-7B`, we set `max_tokens` to 4096, whereas for `Qwen3-4B-Base`, we set it to 32,768. This difference is determined by whether the training data includes LongCoT-style reasoning.

Table 5: Training hyperparameters for `Qwen2.5-0.5B`, `Qwen2.5-7B` and `Qwen3-4B`.

Hyperparameter	Value	Hyperparameter	Value	Hyperparameter	Value
Qwen2.5-0.5B		Qwen2.5-7B		Qwen3-4B	
learning_rate	2.0e-5	learning_rate	5.0e-6	learning_rate	5.0e-5
num_train_epochs	3	num_train_epochs	3	num_train_epochs	3
num_gpus	8	num_gpus	8	num_gpus	8
per_device_train_batch_size	32	per_device_train_batch_size	16	per_device_train_batch_size	2
gradient_accumulation_steps	1	gradient_accumulation_steps	2	gradient_accumulation_steps	2
lr_scheduler_type	cosine	lr_scheduler_type	cosine	lr_scheduler_type	cosine
warmup_ratio	0.1	warmup_ratio	0.1	warmup_ratio	0.1
cutoff_len	4096	cutoff_len	4096	cutoff_len	32768
deepspeed	z0	deepspeed	z2	deepspeed	z2
flash_attn	fa2	flash_attn	fa2	flash_attn	fa2
use_liger_kernel	true	use_liger_kernel	true	use_liger_kernel	true
bf16	true	bf16	true	bf16	true

Table 6: Evaluation hyperparameters.

Hyperparameter	Value
pass@n	n=1
presence_penalty	0.0
frequency_penalty	0.0
repetition_penalty	1.0
temperature	0.0
top_p	1.0
top_k	-1
min_p	0.0
max_tokens	4096 / 32768
min_tokens	0

B Proof of the analytical mixture policy

In this section, we provide a rigorous mathematical derivation for the analytical policy extraction CAUSALMIX-A described in Section 3.3.

Given the estimated state-conditioned marginal return $\hat{\theta}(X_{\text{tar}}) \in \mathbb{R}^K$, our objective is to find the optimal raw mixture strategy T^* that maximizes the expected causal performance gain under the Level-Log formulation. This yields a constrained optimization problem over the probability simplex:

$$\max_{T} \quad \mathcal{J}(T) = \sum_{k=1}^{K} \hat{\theta}_k \log(T_k)$$

$$\text{subject to} \quad \sum_{k=1}^{K} T_k = 1, \quad T_k \geq 0 \quad \forall k \in \{1, \ldots, K\}.$$

To explicitly handle the non-negativity constraints, we reformulate the problem as a minimization problem and apply the Karush–Kuhn–Tucker (KKT) conditions. We minimize $-\mathcal{J}(T)$ and define the Lagrangian $\mathcal{L}(T, \lambda, \mu)$:

$$\mathcal{L}(T, \lambda, \mu) = -\sum_{k=1}^{K} \hat{\theta}_k \log(T_k) + \lambda \left(\sum_{k=1}^{K} T_k - 1 \right) - \sum_{k=1}^{K} \mu_k T_k,$$

where $\lambda \in \mathbb{R}$ is the Lagrange multiplier for the equality constraint, and $\mu_k \geq 0$ are the KKT multipliers for the inequality constraints $T_k \geq 0$.

The KKT optimality conditions require that the optimal solution (T^*, λ^*, μ^*) satisfy: (1) Stationarity: $\frac{\partial \mathcal{L}}{\partial T_k} = -\frac{\hat{\theta}_k}{T_k^*} + \lambda^* - \mu_k^* = 0$, which implies $\lambda^* - \mu_k^* = \frac{\hat{\theta}_k}{T_k^*}$. (2) Primal feasibility: $\sum_{k=1}^{K} T_k^* = 1$ and $T_k^* \geq 0$. (3) Dual feasibility: $\mu_k^* \geq 0$. (4) Complementary slackness: $\mu_k^* T_k^* = 0$.

We analyze the optimal solution by partitioning the domains based on the sign of their estimated marginal returns $\hat{\theta}_k$:

Domains with negative or zero marginal returns ($\hat{\theta}_k \leq 0$). Suppose, for the sake of contradiction, that $T_k^* > 0$. By the complementary slackness condition, $T_k^* > 0 \implies \mu_k^* = 0$. Substituting this into the stationarity condition yields $\lambda^* = \frac{\hat{\theta}_k}{T_k^*}$. Since $\hat{\theta}_k \leq 0$ and $T_k^* > 0$, this implies $\lambda^* \leq 0$. However, there must exist at least one domain j with $\hat{\theta}_j > 0$ and $T_j^* > 0$ (otherwise the objective is unbounded negatively, and empirical mixtures always contain positive-return domains). For that domain j, $\mu_j^* = 0$ implies $\lambda^* = \frac{\hat{\theta}_j}{T_j^*} > 0$, leading to a contradiction. Furthermore, since $\log(T_k) \to -\infty$ as $T_k \to 0$, a negative $\hat{\theta}_k$ pushes the objective value to $+\infty$ as $T_k \to 0^+$. Therefore, the optimal allocation strictly binds at the boundary:

$$T_k^* = 0 \quad \text{for all} \quad \hat{\theta}_k \leq 0.$$

Domains with positive marginal returns ($\hat{\theta}_k > 0$). Let $\mathcal{P} = \{k \mid \hat{\theta}_k > 0\}$ denote the active set. For $k \in \mathcal{P}$, since $T_k^* > 0$ (otherwise the objective drops to $-\infty$), the complementary slackness condition dictates $\mu_k^* = 0$. The stationarity condition simplifies to:

$$\lambda^* = \frac{\hat{\theta}_k}{T_k^*} \implies T_k^* = \frac{\hat{\theta}_k}{\lambda^*}.$$

To determine λ^*, we invoke the primal feasibility condition over the active set $\mathcal{P}$:

$$\sum_{k \in \mathcal{P}} T_k^* = \sum_{k \in \mathcal{P}} \frac{\hat{\theta}_k}{\lambda^*} = 1 \implies \lambda^* = \sum_{k \in \mathcal{P}} \hat{\theta}_k.$$

Substituting λ^* back, we obtain the exact proportional assignment:

$$T_k^* = \frac{\hat{\theta}_k}{\sum_{j \in \mathcal{P}} \hat{\theta}_j} \quad \text{for} \quad k \in \mathcal{P}.$$

By unifying both cases, the global optimal solution maps strictly to zero for non-positive causal effects, and scales proportionally for positive effects. This is analytically identical to applying a Rectified Linear Unit (ReLU) activation to the causal marginal returns followed by L_1 normalization:

$$T_k^{\mathrm{A}} = \frac{[\hat{\theta}_k(X_{\mathrm{tar}})]_+}{\sum_{j=1}^{K} [\hat{\theta}_j(X_{\mathrm{tar}})]_+}, \qquad \text{where} \quad [a]_+ = \max(a, 0).$$

This completes the proof, confirming that our analytical extraction is the mathematically exact closed-form policy under the simplex constraint.