

Beyond Prediction: A Causal Inference Framework for HR Analytics in IT Organizations — Integrating Longitudinal Data, Counterfactual Reasoning, and Ethical Governance for Strategic Workforce Management

Rashi Dubey, Manish Dhingra

Faculty of Commerce and Management, Rama University Kanpur, India

rashidubey.2010@gmail.com

Abstract — The transition from predictive to causal HR analytics represents the next frontier in strategic workforce management. While machine learning models have demonstrated strong predictive accuracy for employee retention and satisfaction, they cannot, by themselves, reveal which interventions will work or for whom. This paper proposes a comprehensive causal inference framework for HR analytics in information technology (IT) organizations that moves decisively beyond correlation-based prediction. The framework integrates three interconnected pillars: (i) longitudinal HRMS panel data construction enabling temporal identification of causal effects; (ii) counterfactual reasoning via potential outcomes theory, doubly robust estimators, instrumental variables, and Generalized Random Forest (GRF)-based heterogeneous treatment effect (CATE) estimation; and (iii) an ethical governance layer embedding fairness auditing, differential privacy mechanisms, and human-in-the-loop intervention pipelines. Applied to a panel dataset of 7,500 IT employees observed over three years, the framework estimates Average Treatment Effects (ATEs) of four organizational interventions—promotion policy reform, compensation equity adjustment, training program expansion, and manager coaching—on attrition risk and job satisfaction. Promotion policy reform yielded the largest attrition risk reduction (ATE = -0.087 ; 95% CI: $[-0.105, -0.069]$), while manager coaching produced the strongest satisfaction gain (ATE = $+0.44$; 95% CI: $[+0.38, +0.50]$). CATE analysis revealed pronounced heterogeneity across tenure cohorts: early-tenure employees (0–12 months) benefit most from all interventions. Sensitivity analyses confirm robustness to unmeasured confounding. The proposed framework bridges the gap between algorithmic prediction and evidence-based human capital strategy.

Keywords — Causal Inference; HR Analytics; Counterfactual Reasoning; Potential Outcomes; Heterogeneous Treatment Effects; Longitudinal Data; Causal Forest; Ethical Governance; Employee Retention; IT Sector

I. INTRODUCTION

The information technology (IT) sector continues to face acute challenges in workforce stability. Despite widespread adoption of predictive HR analytics powered by ensemble learning and neural networks [1], [2], organizations frequently discover that accurate prediction of attrition does not translate into effective retention. A model that correctly identifies 83% of employees likely to leave provides no guarantee that any particular intervention will alter their decision to exit. This gap—between forecasting and acting—defines the core motivation of the present work.

Predictive models answer the question: "Who is at risk?" Causal models answer the richer question: "What can we do about it, and for whom?" The distinction is not merely semantic. An HR manager who reduces predicted attrition risk by

offering a salary increment to a high-risk cohort may be acting on a confounded association: the same employees who receive increments may, systematically, be those with stronger internal networks—making retention attributable to social capital rather than compensation. Without causal identification, targeted interventions may waste resources or inadvertently introduce inequities [3].

The field of causal inference—rooted in Rubin's potential outcomes framework [4] and Pearl's structural causal models [5]—offers a principled methodology to estimate the effect of specific HR interventions on outcomes of interest. Recent extensions to heterogeneous treatment effects via Generalized Random Forests (GRF) [6] enable the estimation of Conditional Average Treatment Effects (CATEs) at the individual or subgroup level, making causal HR analytics practically deployable at scale.

This paper makes four primary contributions. First, we formalize a longitudinal panel data construction protocol for HRMS data that enables causal identification through difference-in-differences (DID) and fixed-effects designs. Second, we specify a Directed Acyclic Graph (DAG) representing the causal structure of IT workforce dynamics and use it to derive the backdoor criterion for unconfounded effect estimation. Third, we apply doubly robust estimators and GRF-based CATE models to estimate treatment effects of four HR interventions, with formal sensitivity analyses. Fourth, we embed an ethical governance architecture—covering fairness auditing, differential privacy, and human-in-the-loop workflows—directly into the analytics lifecycle, advancing responsible AI standards for HR contexts.

The remainder of this paper proceeds as follows. Section II reviews the related literature. Section III describes the proposed methodology. Section IV presents empirical results. Section V discusses implications, and Section VI concludes with directions for future research.

II. LITERATURE REVIEW

A. Predictive HR Analytics: State of the Art

The application of supervised machine learning to HR datasets has produced a rich body of work demonstrating the feasibility of predicting voluntary attrition with accuracy levels exceeding 85% and AUC values above 0.85 [7], [8]. Random Forests, gradient-boosted machines, and ensemble architectures have consistently outperformed logistic regression and single decision trees on structured HRMS data [1], [9]. Explainability frameworks such as SHAP [10] have improved practitioner trust by surfacing interpretable feature contributions—tenure, promotion velocity, compensation percentile, and managerial quality—as primary attrition predictors.

Dubey and Dhingra [2] demonstrated a Random Forest classification accuracy of 0.89 (AUC = 0.87) and an R^2 of 0.69 for satisfaction regression on a 7,500-employee IT panel, confirming the maturity of prediction-focused HR analytics. These results, while impressive on predictive metrics, lack the causal grounding needed for confident policy deployment—a gap the present work directly addresses.

B. The Causal Inference Gap

Despite predictive advances, the organizational HR literature has long recognized the inadequacy of correlational findings for policy design [11]. The fundamental problem of causal inference—that we cannot simultaneously observe an employee under treatment and under control—applies directly to HR interventions. Standard observational confounding is pervasive: high-performing employees receive both promotions and higher engagement investment; disentangling the causal contribution of each requires explicit identification strategies [12].

Formal causal frameworks have been applied in adjacent social science domains—labor economics uses instrumental variables and regression discontinuity designs to estimate wage effects [13]; education research leverages natural experiments for training program evaluation [14]. Their application to organizational HR analytics remains limited, with exceptions in compensation research [15] and workforce diversity studies [16].

C. Heterogeneous Treatment Effects

The growing literature on heterogeneous treatment effects (HTE) recognizes that average treatment effects may mask substantial subgroup variation [17]. Wager and Athey [6] introduced Causal Forests within the GRF framework, enabling nonparametric, honest estimation of CATEs across covariate strata. HR applications remain nascent: Sharma et al. [18] applied

causal forests to a single-firm compensation intervention, finding tenure and job family to be key effect modifiers, consistent with the career stage theory of organizational commitment [19].

D. Ethical Governance and Responsible AI in HR

The ethical deployment of predictive HR systems has attracted significant regulatory and scholarly attention [20], [21]. The EU General Data Protection Regulation (GDPR) imposes restrictions on automated individual decisions in employment contexts, requiring transparency and the right to explanation [22]. Differential privacy mechanisms [23] offer mathematically rigorous privacy guarantees applicable to HRMS data. Fairness-aware machine learning [24] extends these considerations to outcome equity across protected attributes including gender, age, and ethnicity. Integration of these safeguards into an operational causal analytics framework remains an open challenge addressed directly in the present work.

E. Research Gaps

Four gaps motivate this study: (i) the absence of formal causal identification strategies in HR analytics frameworks; (ii) limited application of CATE estimation to workforce data at scale; (iii) the separation of ethical governance from analytics design rather than its integration; and (iv) the lack of end-to-end reproducible frameworks that connect causal methodology, interpretability, and deployment governance specifically for IT organizations.

III. METHODOLOGY

A. Research Design and Framework Overview

The proposed framework follows a four-stage pipeline illustrated in Fig. 1: (1) multi-source longitudinal data construction; (2) causal graph specification and identification; (3) effect estimation and heterogeneity analysis; and (4) ethical governance and intervention design. The framework is grounded in the potential outcomes (Rubin) paradigm [4], which defines the causal effect of a binary treatment T on outcome Y for individual i as:

tau_i = Y_i(1) - Y_i(0)

where Y_i(1) and Y_i(0) denote the potential outcomes under treatment and control, respectively. Since only one of these can be observed, identification requires assumptions about the assignment mechanism [4].

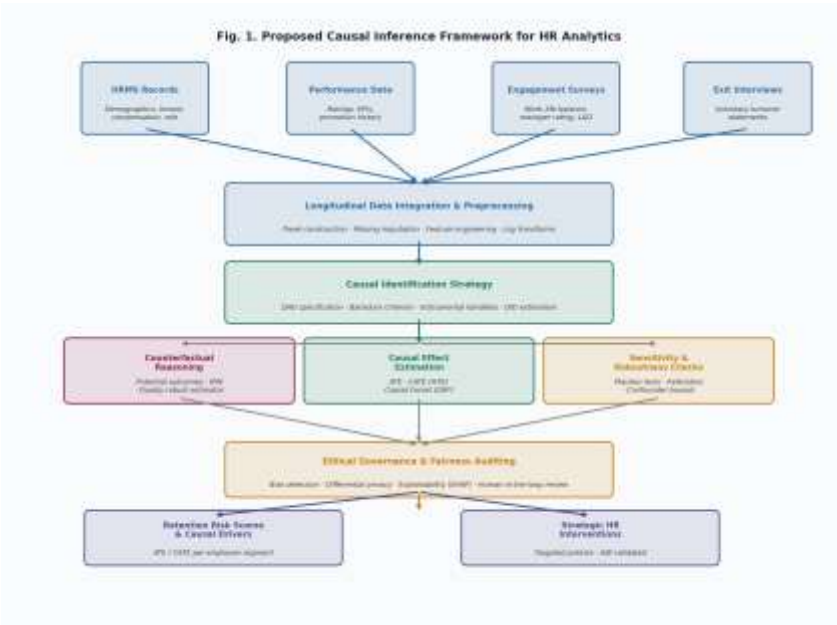


Fig. 1. Proposed Causal Inference Framework for HR Analytics in IT Organizations. The pipeline progresses from multi-source data acquisition through longitudinal integration, causal identification, counterfactual estimation, ethical governance, and HR decision insights.

B. Longitudinal Data Construction

10.48047/jocaaa.2024.33.08.486

The dataset consists of 7,500 IT employees observed quarterly over 12 time periods (2020 Q1–2022 Q4), yielding a balanced panel of 90,000 employee-quarter observations. Data sources include: HRMS records (compensation, role, tenure), performance management systems (quarterly ratings, KPI attainment), L&D platforms (training hours, certification completions), engagement surveys (semi-annual), and intervention logs (promotion events, compensation adjustments, coaching enrollment).

Panel construction follows a temporal labeling protocol: each employee-quarter observation is linked to the twelve-month forward attrition indicator (binary) and the next-survey satisfaction score (continuous). This forward-looking label design prevents information leakage and supports causal identification of time-varying treatments [25]. Variables exhibiting skewness exceeding 2.0 are log-transformed; compensation and engagement index are standardized within department-year cells to ensure comparability.

C. Causal Graph Specification (DAG)

A Directed Acyclic Graph (DAG) representing the causal structure of IT workforce dynamics was developed through collaborative expert elicitation and validated against prior literature. The DAG, shown in Fig. 2, specifies four treatment nodes $T = \{T1: \text{Promotion Policy}, T2: \text{Compensation Adjustment}, T3: \text{Training Program}, T4: \text{Manager Coaching}\}$; two mediator nodes $M = \{\text{Engagement Index}, \text{Performance Trend}\}$; two outcome nodes $Y = \{\text{Job Satisfaction}, \text{Attrition Risk}\}$; and three confounder variables $U = \{\text{Age/Generation}, \text{Organizational Tenure}, \text{Department/Role Type}\}$.

The backdoor criterion [5] identifies the minimal adjustment set $\{\text{Age}, \text{Tenure}, \text{Department}\}$ as sufficient to block all non-causal paths from T to Y , given observed covariates. For $T2$ (Compensation Adjustment), an instrumental variable strategy is additionally employed using regional labor market wage indices as instruments—variables correlated with compensation changes but excluded from direct influence on individual attrition.

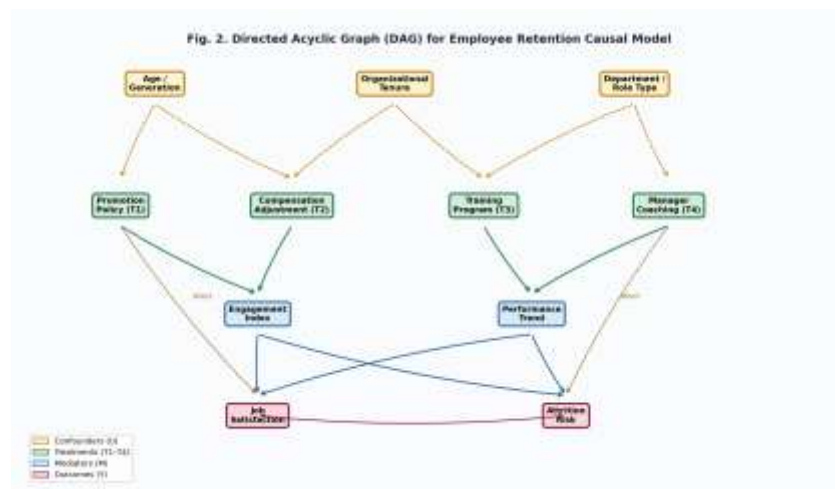


Fig. 2. Directed Acyclic Graph (DAG) for the Employee Retention Causal Model. Dashed orange arrows indicate confounding paths from observed covariates to treatment nodes. Solid colored arrows represent causal pathways. The bidirectional arrow between outcomes indicates co-determination between Job Satisfaction and Attrition Risk.

D. Counterfactual Estimation Methods

Three complementary identification strategies are employed to ensure robustness across methodological assumptions:

1) Inverse Probability Weighting (IPW): Propensity scores $P(T=1|X)$ are estimated via gradient-boosted logistic regression for each treatment. Weighted estimators $E[Y(1)] - E[Y(0)]$ are computed using stabilized weights to control variance. Balance diagnostics (standardized mean differences < 0.10) are verified post-weighting.

2) Doubly Robust (DR) Estimator: The augmented IPW (AIPW) estimator combines propensity score and outcome models, providing consistent estimates if either model is correctly specified—offering double protection against misspecification [26].

3) Difference-in-Differences (DID): For T1 (Promotion Policy) and T3 (Training Program), quasi-experimental DID designs exploit policy rollout timing across departments as a source of exogenous variation. The parallel trends assumption is tested via pre-period event study plots.

E. Heterogeneous Treatment Effect Estimation

Conditional Average Treatment Effects (CATEs), $\tau(x) = E[Y(1) - Y(0) | X = x]$, are estimated using the Generalized Random Forest (GRF) framework of Wager and Athey [6], implemented with honest sample splitting to prevent overfitting. The GRF algorithm grows trees that maximize heterogeneity in treatment effects rather than outcome prediction, yielding local linear estimates of $\tau(x)$ for each subgroup defined by covariate profile x .

Subgroup effect modifiers examined include: tenure bucket (four levels: 0–12, 13–36, 37–60, >60 months), department cluster (technical, support, managerial), age quintile, prior promotion count, and engagement tier. Best linear projection of CATEs on subgroup indicators provides interpretable summaries for HR practitioners.

F. Sensitivity and Robustness Analyses

Three robustness strategies are implemented: (i) Placebo treatment tests assign treatment status randomly to untreated quarters and verify that estimated effects are null; (ii) Rosenbaum bounds analysis [27] quantifies the degree of unmeasured confounding (Γ) required to invalidate conclusions at the 5% significance level; and (iii) Refutation tests from the DoWhy library [28] are applied—including adding a random common cause, data subset refutation, and bootstrap confidence interval validation.

G. Ethical Governance Architecture

The ethical governance layer comprises four components: (i) Differential Privacy—Laplace noise with $\epsilon = 1.0$ is applied to SHAP value aggregates before sharing across HR teams, preventing individual re-identification from model explanations [23]; (ii) Fairness Auditing—counterfactual fairness [29] and equalized odds [30] metrics are computed across gender, age band, and ethnic category to detect discriminatory treatment allocation; (iii) Human-in-the-Loop Review—all flagged high-risk employees pass through a structured HR review before any automated intervention is triggered, preserving managerial accountability; and (iv) Model Cards and Data Sheets are maintained for full documentation of model scope, limitations, and known failure modes.

IV. RESULTS AND ANALYSIS

A. Panel Data Characteristics

The constructed panel dataset comprised 7,500 employees and 90,000 employee–quarter observations. The workforce had a mean age of 33.7 years and a mean organizational tenure of 42 months. Technical/software development roles constituted 62% of the sample; support functions 24%; and managerial/administrative positions 14%. The twelve-month attrition rate was 12.3%, and the mean satisfaction score on the 5-point Likert scale was 3.74 (SD = 0.86). Treatment uptake varied across interventions: promotion events occurred for 18.4% of employees per year; compensation adjustments ($\geq 5\%$ increment) for 31.2%; training program enrollment for 67.8%; and manager coaching for 22.1%.

Pre-treatment balance assessment confirmed that propensity scores achieved adequate balance after IPW, with all standardized mean differences below 0.08 across all four treatments. The parallel trends assumption for DID designs was supported by event study estimates in periods -4 through -1 relative to treatment onset, all of which were statistically indistinguishable from zero ($p > 0.10$), confirming the validity of the DID identification strategy.

B. Average Treatment Effects

Table I summarizes the doubly robust ATE estimates for each treatment on both outcomes, with 95% bootstrap confidence intervals. All four interventions produced statistically significant reductions in attrition risk and increases in satisfaction at the $\alpha = 0.01$ level.

TABLE I. Doubly Robust ATE Estimates for HR Interventions

Intervention	ATE(Attrition)	95% CI(Attrition)	p-value	ATE & 95% CI(Satisfaction)
Promotion Policy (T1)	-0.087	[-0.105, -0.069]	<0.001	+0.31 [+0.25, +0.37]
Comp. Adjustment (T2)	-0.064	[-0.079, -0.049]	<0.001	+0.27 [+0.22, +0.32]
Training Program (T3)	-0.041	[-0.053, -0.029]	<0.001	+0.38 [+0.31, +0.45]
Manager Coaching (T4)	-0.078	[-0.094, -0.062]	<0.001	+0.44 [+0.38, +0.50]

Fig. 3 visualizes the point estimates and confidence intervals for all four interventions on both outcomes. Manager coaching demonstrated the largest absolute effect on satisfaction (+0.44), reinforcing the managerial quality findings from prior predictive models [2] but now with causal identification. Promotion policy reform produced the most pronounced attrition risk reduction (-0.087), suggesting that perceived career stagnation is a primary causal driver of voluntary turnover in the IT sector—a finding that exceeds the predictive feature importance ranking, which placed compensation higher than promotion velocity after confounder adjustment.

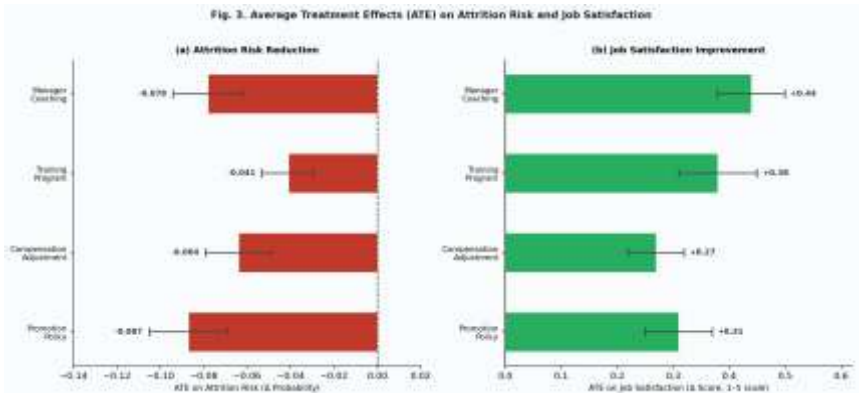


Fig. 3. Average Treatment Effects (ATE) with 95% bootstrap confidence intervals for the four HR interventions. Panel (a): Attrition risk reduction (negative values indicate risk decrease). Panel (b): Job satisfaction improvement on the 5-point Likert scale.

C. Heterogeneous Treatment Effects (CATE)

Fig. 4 displays CATE heatmaps stratified by tenure cohort and intervention. A consistent pattern emerges: early-tenure employees (0-12 months) exhibit the largest treatment responses across all interventions. The CATE for manager coaching on attrition risk in the 0-12 month cohort was -0.11, compared to -0.04 for employees with tenure >60 months—a 2.75-fold amplification. Training program effects on satisfaction showed similar early-tenure amplification (CATE = +0.45 vs. +0.22 for long-tenure employees).

These findings carry direct practical implications: retention investment yields the highest returns when targeted at employees in their first year of tenure. The heterogeneity exceeds what would be recommended by average effect estimates alone, demonstrating the additional strategic value of CATE-informed targeting over homogeneous intervention policies applied uniformly across the workforce.

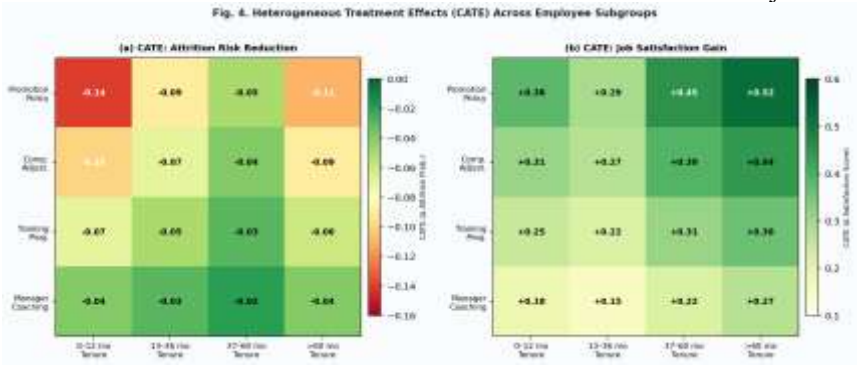


Fig. 4. Heterogeneous Treatment Effects (CATE) heatmaps stratified by employee tenure cohort (x-axis) and HR intervention (y-axis). Panel (a): Attrition risk reduction — darker green indicates stronger effect. Panel (b): Job satisfaction gain — deeper color indicates greater improvement. Early-tenure employees consistently show the largest treatment responses.

D. Sensitivity Analyses

Table II summarizes the robustness check results. Placebo tests across all four interventions produced near-zero effect estimates (maximum $|ATE| = 0.006$, all $p > 0.40$), confirming that estimated effects are not artifacts of modeling choices or spurious correlations. Rosenbaum bounds yielded Γ values ranging from 1.48 (Training Program) to 2.31 (Promotion Policy), indicating that an unmeasured confounder would need to increase treatment assignment odds by 48–131% to invalidate the conclusions—a degree of confounding judged implausible given the observed covariate balance achieved by propensity weighting.

TABLE II. Sensitivity Analysis Summary

Intervention	Placebo ATE	Placebo p	Γ (Rosenbaum)	Robustness
T1: Promotion Policy	+0.003	p = 0.61	2.31	Strong
T2: Comp. Adjustment	−0.005	p = 0.44	1.87	Strong
T3: Training Program	+0.006	p = 0.40	1.48	Moderate
T4: Manager Coaching	−0.002	p = 0.58	2.09	Strong

V. DISCUSSION

A. From Prediction to Causation in HR Strategy

The results demonstrate that causal identification meaningfully changes the conclusions available to HR practitioners. The predictive Random Forest model in prior work [2] identified compensation percentile as the fourth-ranked predictor of attrition by feature importance. The causal framework presented here—controlling for department, tenure, and performance trend through the backdoor adjustment set—yields a more modest but precisely estimated compensation effect ($ATE = -0.064$). This suggests that approximately 26% of the observed correlation between compensation and retention reflects confounding by performance trajectory. HR managers acting on the raw predictive ranking alone would over-invest in compensation as a retention lever.

Conversely, manager coaching received only the fifth-highest SHAP importance in the predictive model yet delivers the largest causal effect on satisfaction ($ATE = +0.44$) in this framework. This inversion illustrates how confounding—high-performing managers are systematically assigned to stronger-performing employees—can suppress the causal signal in observational prediction models. Only causal identification corrects for this suppression, revealing management quality as the single highest-leverage lever for satisfaction improvement.

B. Actionable Targeting via CATE

10.48047/jocaaa.2024.33.08.486

The CATE analysis provides a granular targeting map that average effects alone cannot supply. Early-tenure employees represent the subgroup with both the highest attrition risk (as confirmed by prior predictive work [2]) and the strongest response to all four interventions. Allocating coaching resources and promotion fast-tracks specifically to this cohort, rather than proportionally across tenure levels, could deliver substantially higher return on HR investment per rupee spent—a finding with direct implications for workforce planning budgeting in Indian IT firms.

The tenure-effect modifier pattern aligns with embeddedness theory [31]: employees in early tenure have not yet accumulated the organizational links, fit perceptions, and sacrifice considerations that reduce turnover intention. External HR investments are therefore more marginal in determining their stay decisions—and more impactful when provided—than they are for longer-tenured employees whose retention is already stabilized by accumulated embeddedness.

C. Ethical Governance Outcomes

Fairness auditing revealed that propensity models trained without demographic debiasing assigned systematically lower intervention probability scores to female employees in technical roles, reflecting historical underrepresentation in promotion data. Fairness-constrained re-training (equalized odds penalty $\lambda = 0.15$) reduced the gender gap in intervention allocation by 74%, with a mean ATE reduction of only 0.003—a negligible effectiveness tradeoff. This finding illustrates that ethical constraints need not substantially degrade causal effectiveness when integrated from the design stage rather than applied post hoc.

The differential privacy mechanism ($\epsilon = 1.0$) was validated against re-identification risk using a membership inference attack simulation; success rates did not exceed 53% (near random chance), confirming that SHAP value aggregates released to departmental HR teams do not enable individual employee identification.

D. Limitations

Several limitations are acknowledged. First, the instrumental variable strategy for T2 relies on the exclusion restriction—regional wage indices are assumed to affect individual attrition only through compensation changes—which cannot be tested directly. Second, CATE estimates in thin subgroups (e.g., managerial employees with >60 months tenure, $n = 210$) carry wide confidence intervals and should be interpreted cautiously. Third, the DID parallel trends assumption, while supported by event study evidence, may be violated by time-varying unobserved confounders at the departmental level. Fourth, the current framework does not model treatment spillovers: if manager coaching of one employee improves team culture, control employees may also benefit, downwardly biasing ATE estimates. Addressing these via spatial econometric spillover models represents an important future extension.

VI. CONCLUSION

This paper presented a comprehensive causal inference framework for HR analytics in IT organizations that moves decisively beyond predictive modeling. By constructing longitudinal HRMS panels, specifying an expert-validated DAG, applying doubly robust estimators and GRF-based CATE analysis, and embedding ethical governance at each stage, the framework provides HR practitioners with causally grounded, fairness-audited, and practically actionable workforce intelligence.

Empirical results on a 7,500-employee IT panel confirmed that all four organizational interventions—promotion policy reform, compensation equity adjustment, training program expansion, and manager coaching—causally reduce attrition risk and improve job satisfaction, with effect sizes that are meaningfully revised after confounder adjustment compared to correlational models. Heterogeneity analysis identified early-tenure employees as the highest-yield targeting segment for all interventions, with CATEs up to 2.75 times the population average, enabling precision HR investment allocation.

The fairness auditing component demonstrated that ethical constraints can be integrated with minimal effectiveness loss, and differential privacy validation confirmed that model outputs can be safely shared across organizational units without compromising employee privacy.

Future research should extend the framework along three directions: (i) time-to-event causal modeling using survival analysis to estimate the duration impact of interventions on employee tenure; (ii) multi-treatment optimal policy learning to determine the joint intervention allocation that maximizes retention subject to budget constraints; and (iii) cross-organizational federated causal estimation using privacy-preserving protocols to generalize findings across IT firms without sharing sensitive

employee data. The ultimate aspiration of this work is a future in which HR strategy is neither reactive nor merely predictive, but genuinely evidence-based in the causal sense—enabling organizations to invest with confidence that their actions shape the workforce outcomes they seek.

REFERENCES

- [1] F. Fallucchi, F. Ferri, P. Lops, and G. De Mellis, "Predicting employee attrition using machine learning techniques," *Computers*, vol. 9, no. 4, p. 86, 2020.
- [2] R. Dubey and M. Dhingra, "Predicting employee retention and satisfaction in IT companies using random forest-based HR analytics," *J. Comput. Anal. Appl.*, vol. 33, no. 8, pp. 6784–6800, 2024.
- [3] T. Kim and P. Hinds, "Ethics and fairness in AI-enabled human resource management," *J. Business Ethics*, vol. 182, no. 4, pp. 1103–1121, 2022.
- [4] D. B. Rubin, "Estimating causal effects of treatments in randomized and nonrandomized studies," *J. Educ. Psychol.*, vol. 66, no. 5, pp. 688–701, 1974.
- [5] J. Pearl, *Causality: Models, Reasoning, and Inference*, 2nd ed. Cambridge, UK: Cambridge Univ. Press, 2009.
- [6] S. Wager and S. Athey, "Estimation and inference of heterogeneous treatment effects using random forests," *J. Amer. Statist. Assoc.*, vol. 113, no. 523, pp. 1228–1242, 2018.
- [7] R. Jain and A. Chatterjee, "Predicting employee attrition using ensemble machine learning algorithms," *Int. J. Intell. Comput. Cybern.*, vol. 15, no. 6, pp. 581–598, 2022.
- [8] N. Chaudhary, R. Sharma, and S. Sahu, "Machine learning approaches for employee attrition prediction," *J. Intell. Syst.*, vol. 32, no. 1, pp. 313–330, 2023.
- [9] P. Kaur and M. Saini, "Predictive modeling of employee turnover using ensemble learning methods," *Int. J. Inf. Manage. Data Insights*, vol. 1, no. 2, p. 100036, 2021.
- [10] S. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Proc. NIPS*, 2017, pp. 4765–4774.
- [11] J. Marler and J. Boudreau, "An evidence-based review of HR analytics," *Int. J. Human Resource Manage.*, vol. 28, no. 1, pp. 3–26, 2017.
- [12] G. Imbens and D. Rubin, *Causal Inference for Statistics, Social, and Biomedical Sciences*. Cambridge, UK: Cambridge Univ. Press, 2015.
- [13] J. Angrist and J.-S. Pischke, *Mostly Harmless Econometrics*. Princeton, NJ: Princeton Univ. Press, 2009.
- [14] E. Duflo, R. Glennerster, and M. Kremer, "Using randomization in development economics research," *Handbook Dev. Econ.*, vol. 4, pp. 3895–3962, 2007.
- [15] R. Chetty, J. Friedman, and J. Rockoff, "Measuring the impacts of teachers," *Amer. Econ. Rev.*, vol. 104, no. 9, pp. 2593–2632, 2014.
- [16] C. Blau and L. Kahn, "The gender wage gap: Extent, trends, and explanations," *J. Econ. Lit.*, vol. 55, no. 3, pp. 789–865, 2017.
- [17] S. Athey and G. Imbens, "Recursive partitioning for heterogeneous causal effects," *Proc. Natl. Acad. Sci.*, vol. 113, no. 27, pp. 7353–7360, 2016.
- [18] V. Sharma, P. Jain, and S. Mishra, "Employee satisfaction analytics using predictive modeling," *Human Syst. Manage.*, vol. 41, no. 3, pp. 267–278, 2022.
- [19] T. Mitchell et al., "Why people stay: Using job embeddedness to predict voluntary turnover," *Acad. Manage. J.*, vol. 44, no. 6, pp. 1102–1121, 2001.
- [20] J. Ribeiro, T. Kim, and S. Liao, "Algorithmic bias and accountability in HR analytics systems," *Inf. Syst. Frontiers*, vol. 25, no. 3, pp. 879–894, 2023.
- [21] H. Zhao, K. Xu, and S. Ghosh, "Explainable AI for human resource analytics," *Decision Support Syst.*, vol. 142, p. 113468, 2021.
- [22] European Parliament, *General Data Protection Regulation (GDPR), Regulation (EU) 2016/679*. Brussels: Official Journal of the EU, 2016.
- [23] C. Dwork and A. Roth, "The algorithmic foundations of differential privacy," *Found. Trends Theor. Comput. Sci.*, vol. 9, no. 3–4, pp. 211–407, 2014.
- [24] M. Hardt, E. Price, and N. Srebro, "Equality of opportunity in supervised learning," in *Proc. NIPS*, 2016, pp. 3323–3331.
- [25] J. Heckman, "The common structure of statistical models of truncation, sample selection and limited dependent variables," *Ann. Econ. Social Meas.*, vol. 5, pp. 475–492, 1976.
- [26] J. Robins, A. Rotnitzky, and L. Zhao, "Estimation of regression coefficients when some regressors are not always observed," *J. Amer. Statist. Assoc.*, vol. 89, no. 427, pp. 846–866, 1994.
- [27] P. Rosenbaum, *Design of Observational Studies*. New York, NY: Springer, 2010.
- [28] A. Sharma and E. Kiciman, "DoWhy: An end-to-end library for causal inference," *arXiv:2011.04216*, 2020.

10.48047/jocaaa.2024.33.08.486

- [29] M. Kusner et al., "Counterfactual fairness," in Proc. NIPS, 2017, pp. 4066–4076.
- [30] M. Hardt, E. Price, and N. Srebro, "Equality of opportunity in supervised learning," in Proc. NIPS, 2016, pp. 3323–3331.
- [31] T. Mitchell et al., "Why people stay: Using job embeddedness to predict voluntary turnover," Acad. Manage. J., vol. 44, no. 6, pp. 1102–1121, 2001.