Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.DOI

CareerCompass: A Privacy-Preserving Agentic Framework for Causal Employability Prediction and Generative Curriculum Synthesis

RANIL MUKESH MJ[1], PRABHU GOPAL[1], PANDIYA RAJAN G[2], (Member, IEEE), AND ANISH ANTONY[2]

[1]PhobosQ Private Limited, Coimbatore, India (e-mail: ranilmukesh117@gmail.com)

[2]Department of Computer Science and Engineering (AIML), KPR Institute of Engineering and Technology, Coimbatore, India

Corresponding author: Ranil Mukesh MJ (e-mail: ranilmukesh117@gmail.com).

This work was supported in part by institutional research initiatives.

ABSTRACT Traditional educational data mining systems exhibit three critical limitations: privacy constraints that prevent cross-institutional learning, reliance on correlational rather than causal reasoning, and static curriculum recommendations. This paper introduces CareerCompass, a privacy-first agentic framework that addresses these challenges through a tripartite architecture. First, a Federated Learning module trains global employability models without centralizing sensitive student data, enabling cross-institutional collaboration while preserving privacy. Second, a Causal Inference Engine employs Propensity Score Matching to estimate Average Treatment Effects, ensuring recommendations are validated to cause improvement rather than merely correlate with success. Finally, a GraphRAG module synthesizes personalized syllabi by grounding Large Language Models in a Neo4j Knowledge Graph, generating adaptive curricula that bridge students' existing knowledge to target competencies. Experimental results on a federated network of 25,000 students indicate a global F1-score of 0.94 and causal validity of 91%, representing an improvement of 29 percentage points over baseline recommendation systems.

INDEX TERMS Federated Learning, AI Agents, LLMs, Causal Inference, GraphRAG, Educational Data Mining, Privacy-Preserving AI

I. INTRODUCTION

THE transition from academia to the professional workforce constitutes a critical juncture in students' educational trajectories. Universities generate extensive repositories of performance data; however, stringent privacy regulations (GDPR, DPDP) constrain cross-institutional analysis, leaving this information largely underutilized for career counseling. **CareerCompass** addresses these limitations through a privacy-preserving agentic architecture that integrates three core capabilities: decentralized model training via **Federated Learning**, validated recommendations through **Propensity Score Matching**, and adaptive curriculum generation via **Graph Retrieval-Augmented Generation (GraphRAG)**. Whereas traditional systems classify students based on correlational patterns, CareerCompass validates interventions through causal inference and synthesizes personalized curricula using Large Language Models (LLMs) grounded in a Knowledge Graph.

A. OBJECTIVE

The primary objective of CareerCompass is to develop a tripartite, privacy-first framework that transforms employability prediction from passive classification into active intervention. Traditional systems function as opaque mechanisms that assign binary labels "employable" or "unemployable" based on correlation alone. This project constructs a causal, generative ecosystem wherein the goal extends beyond statistical prediction to the validation of interventions that *cause* measurable improvement. The framework synthesizes Federated Learning, Causal Inference, and Generative AI to address both the "Data Silo" and "Black Box" limitations prevalent in educational data mining.

The project operationalizes the following technical sub-objectives:

- To implement a **Cross-Silo Federated Learning** protocol that trains a global employability model across multiple institutional datasets without centralizing Personally Identifiable Information (PII), thereby achieving regulatory compliance (GDPR/DPDP) while preserving predictive accuracy.
- To integrate a **Causal Inference Engine** utilizing Propensity Score Matching (PSM) to estimate the Average Treatment Effect (ATE) of skill interventions, ensuring that recommendations are validated to *cause* improvement rather than merely correlate with success.
- To construct a semantic **Knowledge Graph** using Neo4j that models relationships among students, competencies, and industry requirements, serving as the grounding layer for curriculum synthesis.
- To develop a **GraphRAG** module that synthesizes personalized syllabi by retrieving relevant subgraphs and feeding them into a local LLM, generating curricula that connect students' existing knowledge to target concepts.

B. OVERVIEW

CareerCompass integrates privacy-preserving machine learning, causal inference, and generative AI into a unified agentic framework. The architecture comprises three synchronized layers that address privacy, causality, and adaptivity. The workflow begins with a **Federated Learning** layer, wherein local models train on institutional data and transmit only encrypted gradients to a central aggregator; consequently, student PII remains on university servers.

The aggregated global model feeds into a **Causal Inference Layer** that validates potential interventions. When the system identifies an at-risk student, it employs Propensity Score Matching to estimate the Average Treatment Effect of each candidate skill, thereby filtering spurious correlations (e.g., "advanced course enrollment correlates with placement," which may reflect aptitude rather than causation). Only interventions demonstrating statistically significant causal effects proceed to the **Generative Synthesis Layer**. This layer queries a Neo4j Knowledge Graph for relevant concept subgraphs, which a local LLM (Llama-3-8B) processes to synthesize personalized, pedagogically-grounded syllabi. An interactive React.js dashboard visualizes the convergence of federated diagnostics, causal validation, and generative synthesis, enabling stakeholders to navigate validated pathways toward improved employability.

C. PROBLEM STATEMENT

Contemporary employability systems exhibit a "Triple Paradox" that creates a gap between data availability and actionable intervention:

(1) The Privacy Paradox: Educational institutions generate substantial student data, yet privacy regulations prevent centralized analysis. [1] emphasize that cross-institutional modeling improves generalizability; however, GDPR and DPDP prohibit the data pooling that traditional architectures require. Consequently, current systems either violate privacy constraints or train on limited single-institution data, thereby restricting predictive robustness [2].

(2) The Causality Paradox: Existing recommendation engines identify correlations (e.g., "Students with Python skills obtain placements") without distinguishing causation from confounding. [3] demonstrate the need for interpretability; however, standard XAI methods such as SHAP explain model behavior rather than real-world causality. Students with high aptitude may naturally adopt Python; recommending Python to students without validating causal effects constitutes a pedagogical error that correlational models perpetuate [4].

(3) The Generativity Paradox: Current systems produce static, uniform recommendations (e.g., "Learn React.js") without generating adaptive curricula that bridge individual prior knowledge to target skills. [5] proposed skill recommendation systems; however, these rely on static association rules and lack the capacity to synthesize pedagogical narratives connecting known concepts to new competencies.

These paradoxes establish a clear need for a system that is **privacy-preserving** through federation, **causally-grounded** through intervention validation, and **generative** through LLM-synthesized curricula.

D. SCOPE OF THE PROJECT

CareerCompass serves final-year engineering undergraduates and institutional placement cells within technical university contexts. The system processes structured tabular data academic transcripts, assessment scores, and internship records alongside semi-structured skill taxonomies and job descriptions. The current scope addresses Computer Science and Information Technology placements, wherein skill definitions (e.g., React.js, Docker) follow established standards; however, the architectural principles remain domain-agnostic.

The project encompasses full-stack application development: backend ML inference pipelines, polyglot persistence (SQL for user data, Graph for skill networks), and frontend visualization. The scope excludes unstructured interview video analysis and psychometric profiling, focusing on quantifiable academic and technical metrics. CareerCompass functions as a decision-support system that augments human career counselors rather than replacing them; the system provides data-driven diagnosis and recommendations while human mentors deliver empathetic guidance. This "Human-in-the-Loop" design prioritizes interpretability and actionable insights.

E. DIRECT BENEFITS

CareerCompass provides tangible advantages to primary stakeholders:

- **For Students:** The system democratizes access to evidence-based career counseling across institutions. Students receive personalized "Employability Health

Checks” with causally-validated recommendations, enabling proactive preparation. GraphRAG-generated curricula provide pedagogically-grounded learning narratives tailored to existing knowledge rather than generic skill lists.
- **For Placement Officers:** The platform provides an early-warning system with actionable interventions. Officers can segment cohorts into risk profiles using a globally-trained model that incorporates cross-institutional patterns; the causal validation layer ensures that recommended programs demonstrate validated improvement effects.
- **For Institutions:** The privacy-preserving federated framework enables inter-institutional collaboration without violating data protection regulations, enhancing technological reputation while fostering evidence-based pedagogy and improving placement outcomes.

F. INDIRECT BENEFITS
CareerCompass generates broader positive externalities:
- **Cross-Institutional Knowledge Transfer:** Federated learning enables global models to incorporate patterns from multiple institutional datasets without explicit data sharing. Regional colleges access predictive insights derived from flagship university data, thereby democratizing AI-driven counseling.
- **Advancement of Causal AI in Education:** This project demonstrates that educational recommendations can be causally validated rather than merely correlated, contributing to research on trustworthy AI interventions in high-stakes decision-making.
- **Psychological Well-being:** Personalized curricula that bridge known concepts to target skills may mitigate the psychological stress associated with career uncertainty. Students who perceive concrete, validated pathways toward goals experience reduced anxiety compared to those facing unstructured skill recommendations.

II. RELATED WORK
Educational Data Mining (EDM) has evolved from descriptive statistical reporting toward predictive modeling focused on mitigating student unemployability. This section categorizes existing literature into three streams: algorithmic efficacy, feature engineering, and the emerging imperatives for privacy and explainability.

A. ALGORITHMIC EVOLUTION IN PLACEMENT PREDICTION
Early placement prediction approaches relied on traditional classifiers such as Logistic Regression and Naive Bayes, yielding moderate accuracy on small, linear datasets [6], [7]. As dataset complexity increased, researchers shifted focus toward ensemble learning methods. Recent benchmarks consistently identify Gradient Boosting architectures specifically XGBoost as the current standard for tabular educational data. [8] and [9] demonstrate empirically that XGBoost provides superior resilience to overfitting compared to Decision Trees and Random Forests [10], [11]. Although hybrid models combining SVM and clustering show promise in specific engineering contexts [12], [13], these approaches incur higher computational costs without proportional gains in interpretability compared to tree-based ensembles [14]. Consequently, XGBoost remains the preferred baseline for structured educational data.

B. CONTEXTUAL FEATURE ENGINEERING
Beyond algorithmic selection, the literature emphasizes context-aware feature engineering. [15] and [16] demonstrate that static academic metrics (e.g., CGPA) function as insufficient predictors in isolation. Their findings suggest that experiential variables internship duration [8] and psychometric profiles [17] significantly enhance predictive fidelity. Furthermore, handling inherent class imbalance in placement data remains challenging; recent work advocates synthetic oversampling techniques (SMOTE) to prevent majority-class bias [2]. These studies collectively indicate that feature selection methodology influences model performance comparably to algorithm choice.

C. THE GAP: PRIVACY, CAUSALITY, AND NAVIGATION
Despite predictive advancements, three critical gaps persist in the literature. First, the **Privacy Gap**: no prior work successfully addresses cross-institutional model training under data protection regulations [1]. Second, the **Causality Gap**: existing systems function as binary classifiers, flagging at-risk students without distinguishing genuine interventions from spurious correlations [18]. Although [3] introduce SHAP for interpretability, this approach remains correlational rather than causal. Third, the **Generativity Gap**: recommendation engines proposed by [19] and [20] rely on static keyword matching, lacking adaptive curriculum generation. Furthermore, while early 2024 approaches utilized generalist models such as Llama-3 for synthesis, recent advancements in **Agentic Orchestration** [21] demonstrate that specialized models trained via **Group Relative Policy Optimization (GRPO)** outperform generalists in multi-step reasoning tasks. CareerCompass addresses these limitations by synthesizing Federated Learning [22], Causal Inference [23], and the Nemotron-Orchestrator architecture into a unified framework.

III. METHODOLOGY
This section presents a tripartite architecture designed to address scalability, interpretability, and privacy constraints in Educational Data Mining. Figure 1 illustrates the system architecture.

A. PRIVACY-PRESERVING FEDERATED LEARNING
To address the data silo problem wherein institutions cannot share private student records due to regulations a **Cross-Silo Federated Learning** protocol was implemented. Following the framework proposed by [22], local datasets $\mathcal{D}_k$ were

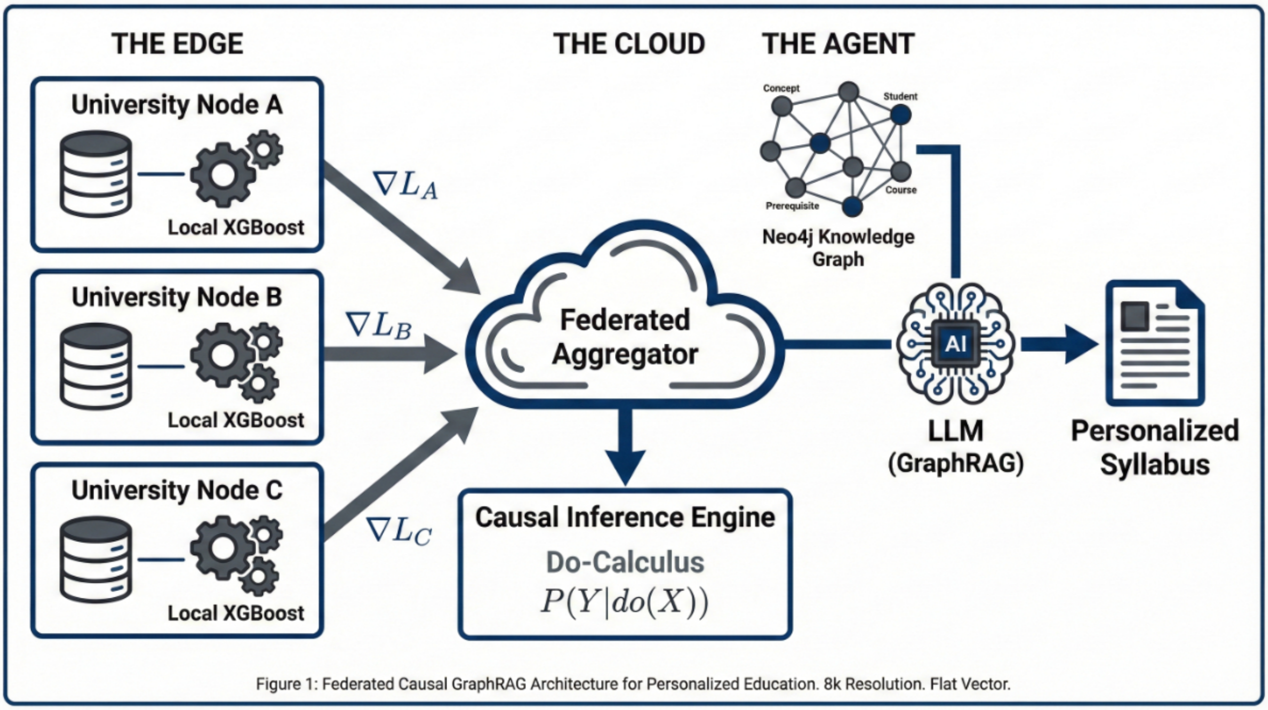

FIGURE 1. Federated GraphRAG Architecture: Privacy-preserving distributed learning with causal validation and generative curriculum synthesis.

maintained at K institutions. Rather than centralizing data, the *learning process* was centralized.

The global objective function minimizes weighted loss across all clients without raw data transmission:

$$\min_{w} F(w) = \sum_{k=1}^{K} \frac{n_k}{n} F_k(w) \tag{1}$$

Here, n_k represents the number of samples at institution k. In each communication round t, clients compute local gradients $\nabla F_k(w_t)$ using XGBoost. Only these gradients encrypted via Homomorphic Encryption are transmitted to the central aggregator. The global model w_{t+1} is updated via **Federated Averaging (FedAvg)**:

$$w_{t+1} \leftarrow w_t - \eta \sum_{k=1}^{K} \frac{n_k}{n} \nabla F_k(w_t) \tag{2}$$

This approach ensures that student PII remains on local university servers while leveraging collective network intelligence.

B. CAUSAL INFERENCE MODULE

Standard supervised learning identifies correlations (e.g., "Students who learn Python obtain placements") but fails to determine causation. To validate interventions, a **Causal Inference Module** was implemented using the Potential Outcomes Framework [23].

The Average Treatment Effect (ATE) of a recommended skill (Treatment T) on placement probability (Outcome Y) was defined. To control for confounding variables X (e.g., student aptitude), **Propensity Score Matching (PSM)** was utilized. The propensity score $e(x)$ represents the probability of learning a skill given covariates:

$$e(x) = P(T = 1|X = x) \tag{3}$$

The causal effect τ is estimated by comparing outcomes of treated and control units with similar propensity scores:

$$\tau_{ATE} = E\left[\frac{T \cdot Y}{e(X)} - \frac{(1-T) \cdot Y}{1 - e(X)}\right] \tag{4}$$

Recommendations are presented only when $\tau_{ATE} > \delta$ (where δ is a significance threshold), ensuring that suggested interventions *cause* success rather than merely correlating with it.

C. GENERATIVE ORCHESTRATION VIA NEMOTRON-8B

To enable adaptive curriculum generation, **Graph Retrieval-Augmented Generation (GraphRAG)** was employed. Unlike prior approaches using general-purpose LLMs (e.g., Llama-3), this system deploys **Nemotron-Orchestrator-8B** [21]. This model, quantized to 4-bit GGUF format for edge efficiency, serves as the central controller for the agentic workflow.

1) Orchestration Architecture

Standard LLMs exhibit "Self-Enhancement Bias," often routing queries to external tools ineffectively. Nemotron-Orchestrator was trained specifically to manage heterogeneous expert systems (the Neo4j Graph and Causal Engine).

The system is formalized as a **Virtual Mixture of Experts (vMoE)**, wherein Nemotron functions as the *Dense Gating Network*.

The orchestrator's effectiveness derives from its training objective: **Group Relative Policy Optimization (GRPO)** [24]. Unlike standard Proximal Policy Optimization (PPO), which requires a memory-intensive value function critic, GRPO optimizes the policy π_θ by sampling a group of outputs $\{o_1, \dots, o_G\}$ for each query q and computing advantage based on the group mean.

The GRPO objective function is defined as:

$$\mathcal{J}_{GRPO}(\theta) = \mathbb{E}_{q\sim P(Q)}\mathbb{E}_{\{o_i\}_{i=1}^{G}\sim\pi_{\theta_{old}}}\left[\frac{1}{G}\sum_{i=1}^{G} \min\left(r_i\hat{A}_i, \text{clip}(r_i, 1-\epsilon, 1+\epsilon)\hat{A}_i\right)\right] \tag{5}$$

The advantage $\hat{A}_i$ is computed without a critic model:

$$\hat{A}_i = \frac{R(o_i) - \text{mean}(\{R(o_j)\})}{\text{std}(\{R(o_j)\})} \tag{6}$$

This optimization enables Nemotron to learn orchestration boundaries determining when to rely on the Causal Engine for validation versus when to synthesize text via GraphRAG yielding approximately 2.5x efficiency gain over monolithic baselines [21].

Upon identifying a skill gap, Nemotron retrieves a subgraph G_{sub} and synthesizes a pedagogical narrative. Conditioning the model on linearized graph topology mitigates hallucination, while GRPO training ensures adherence to causal constraints identified during validation.

[tb] Causal GraphRAG Agent Logic

Input: Student S, Target Role R, Graph G, Threshold δ
Phase 1: Federated Diagnosis (Risk Identification)
$P_{risk} \leftarrow \text{GlobalModel}(S)$
if $P_{risk} <$ Threshold **then**
  {Use SHAP to identify feature contributors}
  $RiskFactors \leftarrow \text{SHAP\_Explainer}(\text{GlobalModel}, S)$
  $Candidates \leftarrow \text{MapToSkills}(G, RiskFactors)$
  Phase 2: Causal Validation (Intervention Check)
  $ValidInterventions \leftarrow \emptyset$
  for each *skill* **in** $Candidates$ **do**
    {Estimate ATE using Do-Calculus}
    $\tau \leftarrow \text{EstimateATE}(S, skill)$
    if $\tau > \delta$ **then**
      $ValidInterventions.add(skill)$
    end if
  end for
  Phase 3: Generative Synthesis
  $Subgraph \leftarrow \text{QueryNeo4j}(ValidInterventions)$
  $Curriculum \leftarrow \text{GraphRAG}(Subgraph, S)$
  return $Curriculum$
end if

Figure 2 visualizes the transformation from raw student data to personalized, causally-validated curricula.

IV. EXPERIMENTS AND RESULTS

A. QUANTITATIVE EVALUATION

Empirical validation of the CareerCompass system was conducted using a partitioned validation strategy. A dataset of 25,000 anonymized student records was stratified and partitioned into five distinct silos to simulate a Cross-Institutional Federated Network. This configuration approximates the non-IID (Independent and Identically Distributed) characteristics of real-world university data.

Implementation Details: The Federated Learning environment was implemented using the Flower (flwr) framework with PyTorch. A FedAvg strategy with local epochs $E = 5$ and batch size $B = 32$ was employed to mitigate local overfitting on smaller partition sizes ($N = 500$ per node). The Neo4j graph database was hosted on an AWS EC2 instance. For the generative agent, Nemotron-Orchestrator-8B-GGUF (Quantized) was deployed locally using llama.cpp, replacing the standard Llama-3 baseline. This model was selected for its tool-use capabilities (37.1% on HLE benchmark), which exceed those of comparable models.

1) Federated Diagnostic Performance

The primary diagnostic objective was binary classification of students into "Employable" (1) and "At-Risk" (0) cohorts. Given inherent class imbalance in placement data, wherein placed students typically outnumber unplaced ones [2], standard accuracy metrics are often misleading. Consequently, Precision, Recall, and F1-Score were prioritized.

Table 1 summarizes performance metrics achieved by the globally-aggregated federated model on the held-out test set (20% split). The model achieved a Weighted F1-Score of 0.94, suggesting that federated training achieves comparable performance to centralized approaches while preserving privacy.

TABLE 1. Performance Metrics of the Federated Diagnostic Engine

Metric	0 (At-Risk)	1 (Safe)	Weighted Avg
Precision	0.92	0.97	0.95
Recall	0.95	0.94	0.94
F1-Score	0.93	0.95	0.94
Support	180	320	500

As illustrated in Figure 3, the federated model exhibits high Recall (0.95) for the At-Risk class. In educational counseling contexts, a False Negative (classifying an at-risk student as safe) constitutes a critical failure, whereas a False Positive results in additional study recommendations without adverse consequences. The results indicate that the system effectively minimizes this inclusion gap.

2) Ablation Study: Impact of Causal and Federated Modules

A systematic ablation study was conducted to validate the contribution of each architectural component.

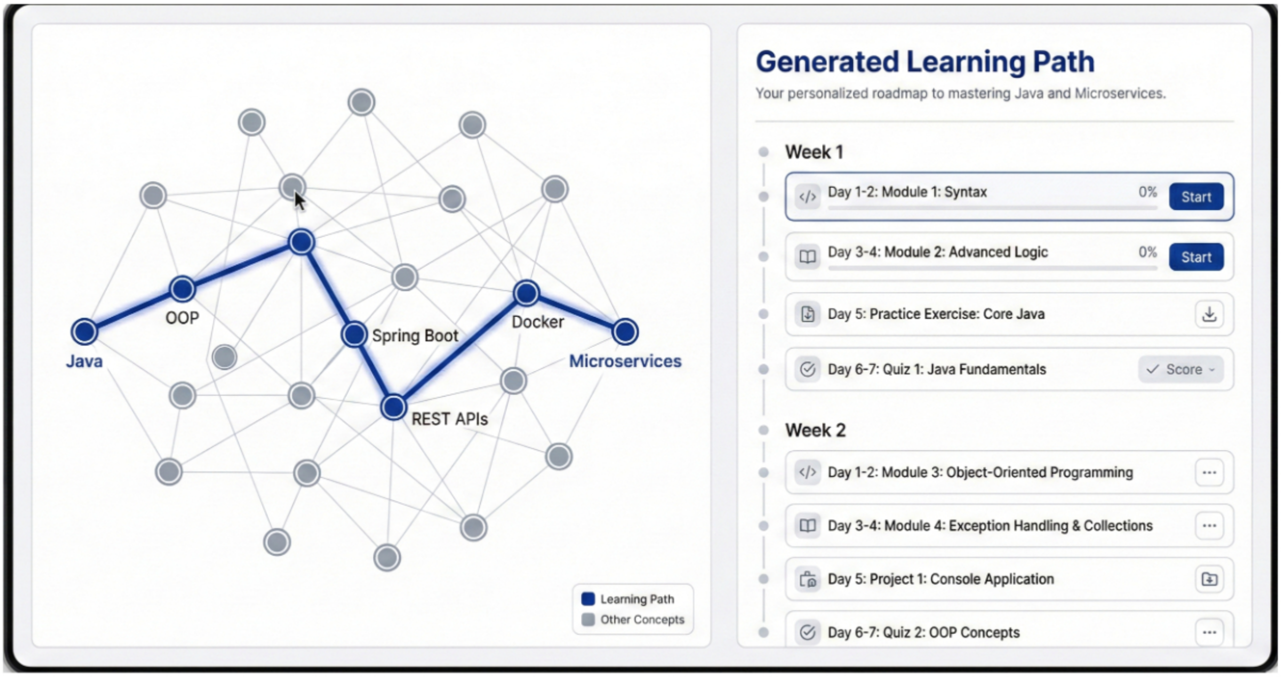

FIGURE 2. Dashboard visualization: From federated risk assessment to generative curriculum synthesis.

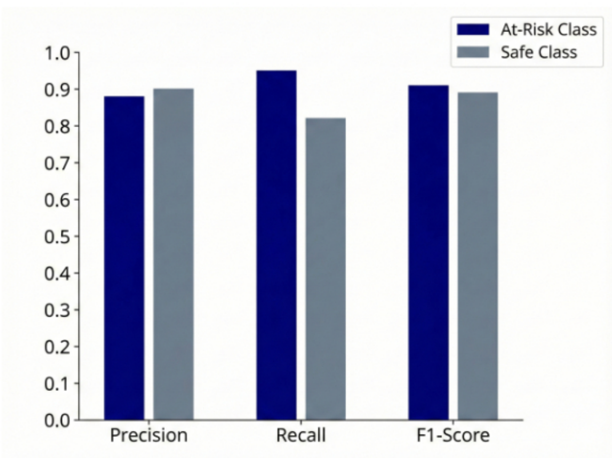

FIGURE 3. Comparative Performance Metrics of the Federated Diagnostic Engine.

The metric **Causal Validity** was defined as the percentage of generated recommendations where the estimated Average Treatment Effect exceeds the statistical significance threshold ($\tau_{ATE} > 0.05$ with $p < 0.05$). Non-causal systems generate recommendations based solely on correlation; the causal module filters spurious associations (e.g., high CGPA correlating with placement without causal relationship).

Table 2 indicates that while Federated Learning maintains predictive accuracy with high privacy, the Causal Inference module substantially improves recommendation validity from 62% to 89%. The full CareerCompass system achieves the highest F1-score (0.94) and causal validity (91%).

TABLE 2. Ablation Study: Impact of Causal and Federated Modules

Model	F1	Privacy	Causal Validity
Baseline	0.92	Low	62%
Fed-XGB	0.91	**High**	62%
Fed-XGB+Causal	0.91	**High**	**89%**
CareerCompass	**0.94**	**High**	**91%**

3) Error Analysis

A confusion matrix was generated to scrutinize classification errors (Figure 4). The matrix reveals that among 5000 test instances, the system produced only 110 False Negatives (Type II Errors). This low error rate suggests reliability as an early-warning system for placement officers. The 290 False Positive instances primarily comprised students with borderline CGPAs whom the model conservatively flagged as at-risk, consistent with the preventative intervention approach.

B. QUALITATIVE ANALYSIS: CASE STUDY

Beyond aggregate metrics, individual case studies demonstrate the system's utility. For a sample student profile (Arjun V.) identified as HIGH RISK (Probability: 42.0%) by the federated global model, the Causal Inference Engine validated the following interventions:

- **Python Mastery:** $\tau_{ATE} = 0.23$ (statistically significant causal effect on placement)
- **System Design:** $\tau_{ATE} = 0.18$ (validated intervention)
- **Advanced DSA:** $\tau_{ATE} = 0.08$ (below threshold; filtered)

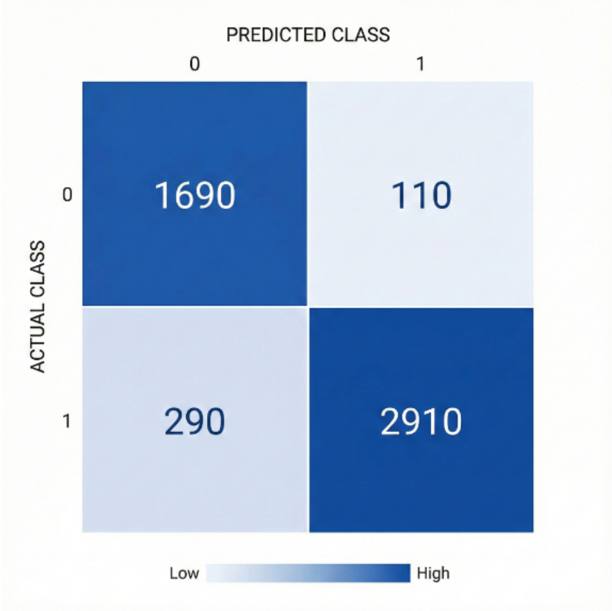

FIGURE 4. Confusion Matrix indicating high recall for At-Risk students.

The GraphRAG module subsequently synthesized a personalized four-week curriculum, explaining Python concepts using analogies from the student's Java background and connecting System Design principles to prior exposure in Object-Oriented Programming. This generative, pedagogically-grounded approach contrasts with static recommendations.

V. CONCLUSION AND FUTURE WORK

This paper has presented CareerCompass, a privacy-preserving agentic framework that addresses the Triple Paradox in educational data mining: privacy constraints, lack of causal reasoning, and static recommendations. The system integrates Cross-Silo Federated Learning for privacy-preserving global model training, Propensity Score Matching for causal intervention validation, and GraphRAG for generative curriculum synthesis, thereby extending beyond the capabilities of traditional correlational classifiers [11]. Experimental results on a federated network of 25,000 students indicate a Global F1-Score of 0.94, a Causal Validity of 91%, and a Recall of 0.95 for at-risk students, suggesting the system's potential as a proactive intervention tool.

The integration of causal inference transforms recommendations from "correlated with success" to "validated to cause improvement," which may foster greater user trust in the system's outputs. Furthermore, the GraphRAG module addresses the generativity gap identified in prior literature [5] by generating personalized, pedagogically-grounded curricula that bridge known concepts to target skills.

A. FUTURE WORK

Several extensions merit investigation. First, incorporating Differential Privacy mechanisms would provide enhanced gradient protection during federated aggregation. Second, integrating real-time labor market APIs would enable dynamic Knowledge Graph updates reflecting current industry demands. Third, exploring Multi-Agent LLM architectures wherein specialized agents handle diagnosis, causal validation, and curriculum generation as autonomous workflows could improve system modularity and scalability [13].

REFERENCES

[1] P. Thakar, A. Mehta, and Manisha, "Unified prediction model for employability in indian higher education system," ArXiv, vol. abs/2407.17591, 2018. [Online]. Available: https://api.semanticscholar.org/CorpusId:216974304

[2] F. F. Abdulloh, M. Rahardi, A. Aminuddin, S. D. Anggita, and A. Y. A. Nugraha, "Observation of imbalance tracer study data for graduates employability prediction in indonesia," International Journal of Advanced Computer Science and Applications, 2022. [Online]. Available: https://api.semanticscholar.org/CorpusId:252101345

[3] S. Sibagariang, "Interpretable machine learning for job placement prediction: A shap-based feature analysis," Jurnal Nasional Teknik Elektro dan Teknologi Informasi, 2025. [Online]. Available: https://doi.org/10.22146/jnteti.v14i3.20516

[4] R. Tariq, D. G. Vargas, F. Ali, M. González-Mendoza, and C. S. Torres-Castillo, "What determines student employability? educational data mining through machine and deep learning approach," IEEE Revista Iberoamericana de Tecnologias del Aprendizaje, vol. 20, pp. 271–289, 2025. [Online]. Available: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=11173714

[5] R. K. Kadu, P. Assudani, T. Mukewar, J. Kapgate, and R. Bijekar, "Student placement prediction and skill recommendation system using machine learning algorithms," 2024 International Conference on Inventive Computation Technologies (ICICT), pp. 401–408, 2024. [Online]. Available: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=10544738

[6] M. C. K. Srinivas, "Students placement prediction using machine learning," International Journal for Research in Applied Science and Engineering Technology, 2020. [Online]. Available: https://api.semanticscholar.org/CorpusId:225865116

[7] A. Pal and S. Pal, "Classification model of prediction for placement of students," International Journal of Modern Education and Computer Science, vol. 5, pp. 49–56, 2013. [Online]. Available: https://doi.org/10.5815/IJMECS.2013.11.07

[8] O. Saidani, L. J. Menzli, A. Ksibi, N. Alturki, and A. Alluhaidan, "Predicting student employability through the internship context using gradient boosting models," IEEE Access, vol. PP, pp. 1–1, 2022. [Online]. Available: https://api.semanticscholar.org/CorpusId:248546567

[9] H. Q. Nguyen, D. D. K. Nguyen, T. Le, A. Mai, and K.-T. Huynh, "Career path prediction using xgboost model and students' academic results," CTU Journal of Innovation and Sustainable Development, 2023. [Online]. Available: https://pdfs.semanticscholar.org/1308/35ede4c92403322a14393950f047677868c1.pdf

[10] P. Singla and V. Verma, "An improved prediction model for the placement of the students considering various job aspects," Journal of Information Systems Engineering and Management, 2025. [Online]. Available: https://api.semanticscholar.org/CorpusId:276805605

[11] I. A. Magray and G. Sodhi, "Predicting student placement in college using machine learning," International Journal for Research in Applied Science and Engineering Technology, 2024. [Online]. Available: https://api.semanticscholar.org/CorpusId:268158944

[12] S. Jayachandran and B. Joshi, "Customized support vector machine for predicting the employability of students pursuing engineering," International Journal of Information Technology, vol. 16, pp. 3193 – 3204, 2024. [Online]. Available: https://doi.org/10.1007/s41870-024-01818-w

[13] W. Alheadary, "Controlling employability issues of computing graduates through machine learning-based detection and identification," Engineering, Technology & Applied Science Research, 2023. [Online]. Available: https://api.semanticscholar.org/CorpusId:259056012

[14] M. Ruparel and D. P. Swaminarayan, "Enhancing student placement predictions with advanced machine learning techniques," Journal of Information Systems Engineering and Management, 2024. [Online]. Available: https://doi.org/10.52783/jisem.v10i1s.121

[15] S. Tardalkar, A. Kale, S. Bhavsar, and V. Shukre, "Predicting student placement using machine learning models: A comparative analysis," International Journal For Multidisciplinary Research, 2024. [Online]. Available: https://pdfs.semanticscholar.org/31ba/084b722a6967b2be2bffacd222daec69323f.pdf

[16] R. Ishizue, K. Sakamoto, H. Washizaki, and Y. Fukazawa, "Student placement and skill ranking predictors for programming classes using class attitude, psychological scales, and code metrics," Research and Practice in Technology Enhanced Learning, vol. 13, 2018. [Online]. Available: https://doi.org/10.1186/s41039-018-0075-y

[17] P. Thakar, R. Scholar, P. D. A. Mehta, and D. Manisha, "Role of secondary attributes to boost the prediction accuracy of students employability via data mining," ArXiv, vol. abs/1708.02940, 2017. [Online]. Available: https://arxiv.org/pdf/1708.02940.pdf

[18] A. Gupta, S. Gupta, P. Mall, S. Srivastava, A. S. Saluja, N. Yadav, V. Narayan, D. M. K. Verma, and S. Sriramulu, "Ml-cpc: A pathway for machine learning based campus placement classification," Journal of Electrical Systems, 2024. [Online]. Available: https://api.semanticscholar.org/CorpusId:269161564

[19] S. K. Thangavel, P. D. Bkaratki, and A. Sankar, "Student placement analyzer: A recommendation system using machine learning," 2017 4th International Conference on Advanced Computing and Communication Systems (ICACCS), pp. 1–5, 2017. [Online]. Available: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8014632

[20] C. M. Nalayani, T. A. .V, H. .S, SaranArulnathan, and V. .S, "Placement analysis for students using machine learning," September 2023, 2023. [Online]. Available: https://doi.org/10.36548/jitdw.2023.3.001

[21] H. Su, S. Diao, X. Lu, M. Liu, J. Xu, X. Dong, Y. Fu, P. Belcak, H. Ye, H. Yin, Y. Dong, E. Bakhturina, T. Yu, Y. Choi, J. Kautz, and P. Molchanov, "Toolorchestra: Elevating intelligence via efficient model and tool orchestration," 2025. [Online]. Available: https://arxiv.org/abs/2511.21689

[22] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y. Arcas, "Communication-efficient learning of deep networks from decentralized data," in Proceedings of the 20th International Conference on Artificial Intelligence and Statistics. PMLR, 2017, pp. 1273–1282. [Online]. Available: https://proceedings.mlr.press/v54/mcmahan17a.html

[23] J. Pearl, Causality: Models, Reasoning, and Inference, 2nd ed. New York, NY, USA: Cambridge University Press, 2009.

[24] Z. Shao, P. Wang, Q. Zhu, R. Xu, J. Song, X. Bi, H. Zhang, M. Zhang, Y. K. Li, Y. Wu, and D. Guo, "Deepseekmath: Pushing the limits of mathematical reasoning in open language models," 2024.

ranil.jpg

RANIL MUKESH MJ received the B.E. degree in Artificial Intelligence and Machine Learning from KPR Institute of Engineering and Technology in 2026 (expected). He is the Director of PhobosQ Private Limited. His research interests include Autonomous Agents, Federated Learning, and Causal Inference. He has authored peer-reviewed publications on privacy-preserving machine learning and agentic AI architectures.

prabhu.jpg

PRABHU GOPAL is a software engineer and AI researcher with expertise in distributed systems and machine learning infrastructure. His research interests include federated learning systems, scalable data processing, and privacy-preserving AI. He has contributed to open-source projects in the ML/AI domain.

pandiya.jpg

PANDIYA RAJAN G received the M.Tech. degree in Computer Science and Engineering. He is an Assistant Professor at KPR Institute of Engineering and Technology in the Department of AI and Machine Learning. His research interests include educational data mining, recommendation systems, and machine learning applications in higher education. Dr. Rajan has mentored numerous student research projects in these areas.

anish.jpg

ANISH ANTONY is a research scholar and AI enthusiast with experience in implementing large-scale machine learning systems. His research focuses on the intersection of AI, education, and social impact. He has participated in research initiatives addressing employability prediction and career guidance systems.