TRÍCH YẾU LUẬN ÁN

a) Tóm tắt mở đầu:

- Tên tác giả: Phạm Thị Tố Nga
- Tên luận án: Một số phương pháp đảm bảo tính công bằng cho các hệ thống học máy trong lĩnh vực giáo dục
- Ngành khoa học của luận án: Khoa học máy tính Mã số: 9480101.01
- Tên đơn vị đào tạo: Trường Đại học Công nghệ, Đại học Quốc gia Hà Nội

b) Nội dung bản trích yếu:

- Mục đích và đối tượng nghiên cứu của luận án: Luận án hướng tới mục tiêu đề xuất các phương pháp đảm bảo tính công bằng cho các hệ thống học máy (ML) trong lĩnh vực giáo dục, nơi dữ liệu thường ở dạng bảng và chứa nhiều thuộc tính nhạy cảm. Trọng tâm nghiên cứu là nâng cao công bằng cho đồng thời nhiều thuộc tính nhạy cảm mà vẫn duy trì hiệu suất mô hình, đồng thời làm rõ mối quan hệ đánh đổi giữa công bằng và hiệu suất nhằm hỗ trợ lựa chọn mô hình tối ưu. Luận án tập trung phát triển ba phương pháp chính: FairEdu (loại bỏ phụ thuộc giữa dữ liệu và thuộc tính nhạy cảm), DPF (cân bằng phân phối bằng dữ liệu tổng hợp), và FairEduPlus (tích hợp hai hướng can thiệp để đảm bảo công bằng hai chiều). Ngoài ra, chỉ số đánh đổi giữa công bằng và hiệu suất cũng được đề xuất nhằm định lượng mối quan hệ này một cách hệ thống. Đối tượng nghiên cứu là các mô hình học máy ứng dụng trong giáo dục, bao gồm Logistic Regression, Decision Tree, Random Forest, Gradient Boosting và Neural Network cơ bản, được huấn luyện trên dữ liệu chứa nhiều thuộc tính nhạy cảm như giới tính, chủng tộc, tuổi, sức khỏe, tình trạng nợ và nơi ở.
- Phương pháp nghiên cứu: Luận án kết hợp phương pháp nghiên cứu lý thuyết và triển khai thực nghiệm các kết quả thu được so với các kết quả mới nhất trên thế giới. Dây là phương pháp nghiên cứu hiện đại, có độ tin cậy cao và đang được sử dụng phổ biến.
- Các kết quả chính và kết luận: Luận án đã nghiên cứu và phát triển các phương pháp đảm bảo tính công bằng cho các mô hình học máy trong giáo dục, nơi dữ liệu thường dạng bảng, chứa nhiều thuộc tính nhạy cảm và mất cân bằng. Bốn kết quả chính đạt được gồm: (1) Phương pháp FairEdu – kỹ thuật tiền xử lý dựa trên hồi quy đa biến, loại bỏ sự phụ thuộc giữa thuộc tính nhạy cảm và dữ liệu huấn luyện, giúp cải thiện công bằng khi xét đồng thời nhiều thuộc tính. (2) Phương pháp DPF (Data Partitioning Fairness) – sử dụng kỹ thuật sinh dữ liệu tổng hợp (SDG) để cân bằng phân phối giữa các nhóm giao thoa, tăng tính đại diện của dữ liệu. (3) Phương pháp FairEduPlus – tích hợp FairEdu và DPF thành cơ chế can thiệp hai chiều, vừa hiệu chỉnh dữ liệu, vừa cân bằng phân phối, đảm bảo công bằng toàn diện mà vẫn duy trì hiệu suất ổn định. (4) Chỉ số đánh đổi – công cụ định lượng mối quan hệ giữa công bằng và hiệu suất mô hình. Kết quả thực nghiệm trên bốn bộ dữ liệu giáo dục cho thấy các phương pháp đề xuất giảm rõ rệt chênh lệch công bằng (DI, SPD, AOD, EOD) trong khi vẫn giữ hiệu suất ở mức cao. Về ý nghĩa, luận án đóng góp

một khung phương pháp luận mới cho nghiên cứu công bằng trong học máy, có khả năng mở rộng sang các lĩnh vực nhạy cảm khác như y tế hoặc tuyển dụng. Tuy nhiên, các phương pháp hiện mới được kiểm chứng trên dữ liệu dạng bảng và cần được tự động hóa để áp dụng rộng rãi hơn trong tương lai.

Hà Nội, ngày tháng năm 2025

NGHIÊN CỨU SINH	**CÁN BỘ HƯỚNG DẪN**
(*Ký và ghi rõ họ, tên*)	(*Ký và ghi rõ họ, tên*)
Phạm Thị Tố Nga	**Phạm Ngọc Hùng**

XÁC NHẬN CỦA CƠ SỞ ĐÀO TẠO/
CONFIRMATION FROM THE TRAINING UNIVERSITY

THE ABSTRACT OF DOCTORAL THESIS

a) Preliminary Information

- Author: Pham Thi To Nga
- Title of the doctoral thesis: Several Methods for Ensuring Fairness in Machine Learning Systems in Education
- Major: Computer Science
- Code: 9480101.01
- Training Institution: University of Engineering and Technology, Vietnam National University, Hanoi

b) Content Summary

- Research Objectives and Scope: The dissertation aims to propose effective methods for ensuring fairness in machine learning (ML) systems applied to educational data, which are typically tabular and contain multiple sensitive features. The focus is to enhance fairness simultaneously across several sensitive features while maintaining model performance, and to clarify the trade-off relationship between fairness and accuracy to support optimal model selection. Three main methods are developed: (1) FairEdu – a preprocessing method that removes dependencies between sensitive attributes and non-sensitive features; (2) DPF (Data Partitioning Fairness) – a data-level balancing method that uses synthetic data generation (SDG) to equalize the representation of intersectional groups; and (3) FairEduPlus – a hybrid framework integrating both approaches to achieve two-dimensional fairness (distributional and relational). In addition, (4) a trade-off index is proposed to systematically quantify the balance between fairness and predictive performance. The research objects are ML models commonly used in educational applications, including Logistic Regression, Decision Tree, Random Forest, Gradient Boosting, and basic Neural Networks, trained on datasets containing multiple sensitive variables such as gender, ethnicity, age, health status, debt condition, and residence.

- Research Methods: The dissertation combines theoretical analysis with empirical experiments, comparing the proposed methods against the most recent state-of-the-art approaches worldwide. This is a modern and reliable research methodology, widely adopted in contemporary AI and data science studies.

- Main Results and Conclusions: The dissertation investigates and develops fairness-enhancing methods for machine learning models applied in education, where data are tabular, imbalanced, and contain multiple sensitive features. Four major results are achieved: (1) FairEdu Method: A multivariate regression-based preprocessing technique that removes statistical dependencies between sensitive and non-sensitive attributes, effectively improving fairness across multiple sensitive variables. (2) DPF (Data Partitioning Fairness): A method leveraging synthetic data generation (SDG) to rebalance the distribution among intersectional subgroups, thereby increasing the representativeness

of the training data. (3) FairEduPlus Method: A comprehensive, two-dimensional approach integrating FairEdu and DPF to simultaneously adjust data features and group distributions, achieving overall fairness without significant performance loss. (4) Trade-off Index: A new quantitative measure that evaluates the balance between model fairness and predictive accuracy. Experiments on four educational datasets demonstrate that the proposed methods substantially reduce fairness disparities (measured by DI, SPD, AOD, and EOD) while maintaining high model performance. In terms of contribution, the dissertation introduces a systematic methodological framework for fairness assessment in ML, with potential extensions to other sensitive domains such as healthcare and recruitment. However, the proposed methods have been validated primarily on tabular data and still require further automation to be applied widely in large-scale, real-world systems.

Hanoi, October 15, 2025

PHD STUDENT	**SUPERVISOR**
(Signed and full name)	*(Signed and full name)*
Pham Thi To Nga	**Pham Ngoc Hung**

CONFIRMATION FROM THE TRAINING UNIVERSITY