

THÔNG TIN VỀ LUẬN ÁN TIẾN SĨ

1. Họ và tên nghiên cứu sinh: Phạm Thị Tố Nga 2. Giới tính: Nữ
3. Ngày sinh: 27/10/1978..... 4. Nơi sinh: Phú Thọ
5. Quyết định công nhận nghiên cứu sinh số 25/QĐ-CTSV ngày 14 tháng 1 năm 2022 của Hiệu trưởng Trường Đại học Công nghệ, Đại học Quốc gia Hà Nội.
6. Các thay đổi trong quá trình đào tạo: Không
(ghi các hình thức thay đổi và thời gian tương ứng)
7. Tên đề tài luận án: Một số phương pháp đảm bảo tính công bằng cho các hệ thống học máy trong lĩnh vực giáo dục
(tên luận án chính thức đề nghị đánh giá luận án)
8. Ngành đào tạo: Khoa học máy tính 9. Mã số: 9480101
10. Cán bộ hướng dẫn khoa học: (ghi rõ chức danh khoa học, học vị, họ và tên)
 - Hướng dẫn chính: GS.TS Nguyễn Đức Anh - Trường Đại học Đông Nam Na Uy
 - Hướng dẫn phụ: PGS.TS. Phạm Ngọc Hùng - Trường Đại học Công nghệ - ĐHQGHN
11. Tóm tắt các **kết quả mới** của luận án:
 - 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. Đây 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, từ đó nâng cao tính công bằng cho các mô hình học máy ngay từ khâu tiền xử lý. (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.

12. Khả năng ứng dụng trong thực tiễn: *(nếu có)*

Các phương pháp đề xuất như FairEdu, DPF, và FairEduPlus đều có thể trực tiếp áp dụng cho các mô hình học máy trong lĩnh vực giáo dục có sử dụng dữ liệu đào tạo dạng bảng nhằm kiểm tra và nâng cao tính công bằng cho hệ thống.

13. Những hướng nghiên cứu tiếp theo: *(nếu có)*

Trên cơ sở các kết quả đạt được, luận án đề xuất một số định hướng nghiên cứu tiếp theo nhằm nâng cao hiệu quả và mở rộng khả năng ứng dụng trong thực tiễn, bao gồm:

- Mở rộng phạm vi dữ liệu và mô hình, bao gồm các lĩnh vực ngoài giáo dục (y tế, tài chính) và các mô hình hiện đại như deep learning, Transformer, nhằm đánh giá khả năng tổng quát và mở rộng của phương pháp.
- Phát triển các thước đo công bằng toàn diện hơn, đặc biệt cho trường hợp đa thuộc tính nhạy cảm, đồng thời hoàn thiện chỉ số đánh đổi để hỗ trợ lựa chọn mô hình trong thực tiễn.
- Nâng cao kỹ thuật sinh dữ liệu tổng hợp cho các tình huống mất cân bằng nghiêm trọng hoặc dữ liệu thưa, nhằm cải thiện đại diện cho các nhóm giao thoa.
- Xây dựng khung công bằng tự động, có khả năng phát hiện, điều chỉnh và đánh giá công bằng trên nhiều loại dữ liệu và mô hình, giảm phụ thuộc vào can thiệp thủ công.

14. Các công trình đã công bố có liên quan đến luận án:

1. Pham, N., Pham-Ngoc, H., Nguyen-Duc, A. (2023). *Fairness Requirement in AI Engineering – A Review on Current Research and Future Directions*. In: Gupta, V., Rubalcaba, L., Gupta, C., Hanne, T. (eds) Sustainability in Software Engineering and Business Information Management. SSEBIM 2022. Lecture Notes in Information Systems and Organisation, vol 62. Springer, Cham. https://doi.org/10.1007/978-3-031-32436-9_1.
2. N. Pham, P. N. Hung, and A. Nguyen-Duc, *Fairness for machine learning software in education: A systematic mapping study*, J. Syst. Softw., p. 112244, Oct. 2024, doi: 10.1016/j.jss.2024.112244 (Q1)
3. N. Pham, M. K. Do, T. V. Dai, P. N. Hung, and A. Nguyen-Duc, *FAIREDU: A Multiple Regression-Based Method for Enhancing Fairness in Machine Learning Models for Educational Applications*, Expert Syst. Appl., Dec. 2024, doi.org/10.1016/j.eswa.2024.126219 (Q1)
4. N. Pham, Minh Kha Do, Quang Trung Doan, Pekka Abrahamsson, Anh Nguyen-Duc, Pham Ngoc Hung, *FaireduPlus: Enhancing Intersectional Fairness in Education Focused Machine Learning Using Synthetic Data*, submitted to IST (Q1)

Ngày tháng năm 20
Xác nhận của cán bộ hướng dẫn
(Kí và ghi rõ họ tên)

Ngày 24 tháng 3 năm 2026
Nghiên cứu sinh
(Kí và ghi rõ họ tên)

INFORMATION ON DOCTORAL THESIS

1. Full name : Pham Thi To Nga 2. Sex: Female
3. Date of birth: 27/10/1978..... 4. Place of birth: Phu Tho
5. Admission decision number: 25/QĐ-CTSV. Dated: January 14, 2022
6. Changes in academic process: Non
(List the forms of change and corresponding times)
7. Official thesis title: Several Methods for Ensuring Fairness in Machine Learning Systems in Education
8. Major: Computer Science 9. Code: 9480101
10. Supervisors:
 1. Prof. Dr. Nguyen Duc Anh – University of South-Eastern Norway
 2. Assoc. Prof. Dr. Pham Ngoc Hung – University of Engineering and Technology, VNU
11. Summary of the **new findings** of the thesis:
 - 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, and consequently enhancing fairness in machine learning systems from the pre-processing stage. (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.

12. Practical applicability, if any:

The proposed methods, including FairEdu, DPF, and FairEduPlus, can be directly applied to machine learning models in the educational domain that use tabular training data to evaluate and enhance the fairness of the system.

13. Further research directions, if any:

Based on the achieved results, the dissertation proposes several future research directions to enhance effectiveness and expand practical applicability, including:

- Extending the scope of datasets and models, including domains beyond education (e.g., healthcare, finance) and modern architectures such as deep learning and Transformer-based models, to evaluate generalizability and scalability.
- Developing more comprehensive fairness metrics, especially for multi-sensitive attribute settings, and further refining the trade-off index to support practical model selection.

- Improving synthetic data generation techniques for highly imbalanced or sparse data scenarios, to enhance representation of intersectional subgroups.
- Designing an automated fairness framework capable of detecting, mitigating, and evaluating fairness across various data types and models, thereby reducing reliance on manual intervention.

14. Thesis-related publications:

1. Pham, N., Pham-Ngoc, H., Nguyen-Duc, A. (2023). *Fairness Requirement in AI Engineering – A Review on Current Research and Future Directions*. In: Gupta, V., Rubalcaba, L., Gupta, C., Hanne, T. (eds) Sustainability in Software Engineering and Business Information Management. SSEBIM 2022. Lecture Notes in Information Systems and Organisation, vol 62. Springer, Cham. https://doi.org/10.1007/978-3-031-32436-9_1.
2. N. Pham, P. N. Hung, and A. Nguyen-Duc, *Fairness for machine learning software in education: A systematic mapping study*, J. Syst. Softw., p. 112244, Oct. 2024, doi: 10.1016/j.jss.2024.112244 (Q1)
3. N. Pham, M. K. Do, T. V. Dai, P. N. Hung, and A. Nguyen-Duc, *FAIREDU: A Multiple Regression-Based Method for Enhancing Fairness in Machine Learning Models for Educational Applications*, Expert Syst. Appl., Dec. 2024, doi.org/10.1016/j.eswa.2024.126219 (Q1)
4. N. Pham, Minh Kha Do, Quang Trung Doan, Pekka Abrahamsson, Anh Nguyen-Duc, Pham Ngoc Hung, *FaireduPlus: Enhancing Intersectional Fairness in Education Focused Machine Learning Using Synthetic Data*, submitted to IST (Q1)

Date:

Date:

Signature:

Signature:

Full name: Phạm Ngọc Hùng

Full name: Phạm Thị Tố Nga