

# THÔNG TIN CHUNG CỦA NHÓM

- Link YouTube video của báo cáo (tối đa 5 phút):  
<https://www.youtube.com/watch?v=06ijXp704Nk>
- Link slides (dạng .pdf đặt trên Github của nhóm):  
[https://github.com/khanhtrphuocbao/Research-Methodology/blob/main/Khanh  
Truong Phuoc Bao - CS2205.SEP2025.DeCuong.FinalReport.Template.Slide.pdf](https://github.com/khanhtrphuocbao/Research-Methodology/blob/main/Khanh%20Truong%20Phuoc%20Bao%20-%20CS2205.SEP2025.DeCuong.FinalReport.Template.Slide.pdf)
- Mỗi thành viên của nhóm điền thông tin vào một dòng theo mẫu bên dưới
- Sau đó điền vào Đề cương nghiên cứu (tối đa 5 trang), rồi chọn Turn in
- *Lớp Cao học, mỗi nhóm một thành viên*

- Họ và Tên: Trương Phước  
Bảo Khanh
- MSHV: 250101033



- Lớp: CS2205.RM
- Tự đánh giá (điểm tổng kết môn): 9.5/10
- Số buổi vắng: 0
- Số câu hỏi QT cá nhân: 7
- Số câu hỏi QT của cả nhóm: 7
- Link Github:  
[https://github.com/khanhtrphuocbao/Research-  
Methodology](https://github.com/khanhtrphuocbao/Research-Methodology)

# ĐỀ CƯƠNG NGHIÊN CỨU

## TÊN ĐỀ TÀI (IN HOA)

PHÂN LOẠI VÀ XÁC ĐỊNH TỰ ĐỘNG YẾU TỐ BẤT THƯỜNG TRONG TIN TUYỂN DỤNG TIẾNG VIỆT

## TÊN ĐỀ TÀI TIẾNG ANH (IN HOA)

CLASSIFYING AND DETECTING ABNORMAL ASPECTS IN JOB RECRUITMENT IN VIETNAMESE

## TÓM TẮT *(Tối đa 400 từ)*

Trong nghiên cứu này, qua quá trình thu thập và gán nhãn dữ liệu, nhóm chúng tôi đã thành công xây dựng bộ dữ liệu gồm 12.054 mẫu tin tuyển dụng tiếng Việt đã được gán nhãn đầy đủ để phục vụ cho ba tác vụ chính của bài toán là phân lớp mức độ khía cạnh, phân lớp nhãn chung và sinh câu giải thích. Ở hai tác vụ phân lớp đầu tiên, chúng tôi sử dụng các mô hình tiền huấn luyện đa ngôn ngữ như XLM-RoBERTa, DistilBERT và các mô hình đơn ngôn ngữ được huấn luyện trên ngôn ngữ tiếng Việt như PhoBERT, ViSoBERT, CafeBERT thu được các véc-tơ ngữ nghĩa, làm tiền đề cho việc phân lớp phía sau. Việc xây dựng các kiến trúc phân lớp cũng mang ý nghĩa quan trọng. Nhóm chọn ra 2 kiến trúc phổ biến và mang lại hiệu suất khá tốt trên các dữ liệu ngôn ngữ là LSTM và CNN để xây dựng đuôi phân lớp của mô hình. Việc sử dụng LSTM và CNN mỗi loại riêng biệt với feature từ các mô hình pre-trained mang lại những lợi ích riêng của từng mạng. LSTM tập trung vào việc mô hình hóa mối quan hệ ngữ nghĩa dài hạn và xử lý chuỗi, trong khi CNN phát huy sức mạnh trong việc trích xuất đặc trưng cục bộ và mô hình hóa mối quan hệ không gian giữa các từ. Lựa chọn giữa LSTM và CNN (hoặc cả hai) phụ thuộc vào bối cảnh cụ thể của vấn đề và mục tiêu đặt ra trong tác vụ phân lớp dữ liệu văn bản. Đối với tác vụ thứ ba, nhóm sử dụng các mô hình ngôn ngữ tiền huấn luyện như BARTpho và ViT5.

## GIỚI THIỆU *(Tối đa 1 trang A4)*

Sự phát triển của các trang mạng xã hội đã mở ra một cơ hội lớn công việc mới và đa dạng cho việc tuyển dụng trực tuyến. Cùng với sự phát triển này, số lượng các tin tuyển dụng không có thật hoặc có khả năng lừa đảo cũng tăng lên đáng kể. Việc một người tìm kiếm việc làm trực tuyến gặp phải các tin tuyển dụng giả mạo có mô tả hấp dẫn như mức lương cao, giờ giấc làm việc linh động, cơ hội làm việc từ xa và phát triển nghề nghiệp ngày càng phổ biến. Đặc biệt trên các trang web tổng hợp mua bán và tìm việc làm, sự xuất hiện của những tin tuyển dụng không có thật, có khả năng lừa đảo hoặc cung cấp các thông tin gây hiểu lầm về công việc đã trở thành một vấn đề

nghiêm trọng gây ra nhiều rủi ro cho người tìm việc làm ở Việt Nam, lôi kéo người tìm việc làm để lợi dụng hoặc chiếm đoạt thông tin cá nhân, tài sản của người lao động. Hơn nữa, các công ty hoặc tổ chức khác đều có khả năng bị giả mạo thông tin. Điều này gây ảnh hưởng không nhỏ đến hình ảnh và độ uy tín của các công ty nói trên.

Chính vì thế, nhằm ngăn chặn kẻ xấu sử dụng việc tuyển dụng để lừa đảo và bảo vệ quyền lợi của các bên tham gia, đóng góp của chúng tôi đối với bài toán “Warning Job Advertising Detection”

### **MỤC TIÊU** (*Viết trong vòng 3 mục tiêu*)

Mục tiêu chính của đề tài là xây dựng một bộ dữ liệu nhận biết các tin tuyển dụng bất thường ở mức độ khía cạnh giúp nhận biết và ngăn chặn các tin tuyển dụng có khả năng lừa đảo trên các cổng thông tin tìm việc làm ở Việt Nam. Điều này nhằm bảo vệ quyền lợi của người tìm việc và người tuyển dụng, cũng như cung cấp các hỗ trợ để đưa ra quyết định chính xác hơn. Gồm có ba mục tiêu nhỏ hơn:

1. Xây dựng một quy trình gán nhãn cụ thể cho bài toán.
2. Xây dựng bộ dữ liệu về các tin tuyển dụng bất thường, có khả năng lừa đảo cho tiếng Việt.
3. Xây dựng các mô hình baseline và phân tích lỗi trên bộ dữ liệu.

### **NỘI DUNG VÀ PHƯƠNG PHÁP**

#### **Nội dung 1: Các công trình nghiên cứu liên quan**

##### **1.1. Bài toán phát hiện tuyển dụng lừa đảo (Online Recruitment Fraud - ORF)**

Lĩnh vực này còn khá mới mẻ, đặc biệt là tại Việt Nam, dù tình trạng lừa đảo qua mạng ngày càng nghiêm trọng. Bài toán ORF được Vidros [1] và cộng sự đề xuất lần đầu vào năm 2016, chỉ ra các quy luật thực nghiệm của lừa đảo như tương tác ngắn hoặc thiếu giao thức giao tiếp chung. Năm 2017, nhóm này công bố bộ dữ liệu EMSCAD [2] (Employment Scam Aegean Dataset) gồm ~17.880 tin tuyển dụng để phục vụ nghiên cứu. Tại Việt Nam hiện chưa có bộ dữ liệu ORF chuyên biệt. Các nghiên cứu hiện có chủ yếu tập trung vào phân loại tiêu đề/mô tả công việc ngành IT sử dụng Deep Learning (TextCNN, Bi-GRU-LSTM-CNN) [4].

##### **1.2. Phân tích cảm xúc dựa trên khía cạnh (Aspect-based Sentiment Analysis - ABSA)**

Định nghĩa: ABSA xác định cảm xúc (tích cực/tiêu cực/trung lập) theo từng khía cạnh cụ thể của văn bản.

Bối cảnh: Các phương pháp hiện tại chủ yếu dựa trên dữ liệu cũ và mang đặc trưng văn hóa riêng, nguồn dữ liệu tiếng Việt rất khan hiếm.

Chúng tôi áp dụng ý tưởng của ABSA nhưng thay đổi mục tiêu. Thay vì phân tích cảm xúc, nhóm tập trung đánh giá mức độ xác thực và tính nhất quán của các nhóm

thuộc tính (khía cạnh) trong tin tuyển dụng, dựa trên ý tưởng từ nghiên cứu của Văn Thìn và cộng sự [3].

### **1.3. Explainable AI (AI giải thích được)**

Mục tiêu: Giúp con người hiểu được lý do tại sao hệ thống AI đưa ra một quyết định cụ thể. Các nghiên cứu trước đây như bài toán QA hay kiểm chứng thông tin đa phương thức đã đề xuất việc kèm theo các câu giải thích (bằng ngôn ngữ tự nhiên) cho kết quả dự đoán. Ở đây, chúng tôi tiến hành mở rộng bộ dữ liệu bằng cách thêm đầu ra thứ ba là câu giải thích bằng tiếng Việt, nhằm làm rõ lý do tại sao một tin tuyển dụng lại được gán nhãn là cảnh báo hay bình thường.

#### **Nội dung 2: Xác định bài toán**

Thông thường, trên các trang tìm việc làm, nhà tuyển dụng khi đăng tải tin tuyển dụng thường chia các thông tin này thành các mục nhỏ để người tìm việc dễ dàng sàng lọc thông tin. Các mục thông tin này có thể là yêu cầu công việc, quyền lợi công việc, vv. Đặc điểm chung của các thuộc tính thành phần bên trong từng mục là luôn có liên quan đến nhau, gồm tất cả 5 khía cạnh của tin tuyển dụng bao gồm: title, description, company, poster, other và các thuộc tính bên trong.

Về định nghĩa, bài toán được chúng tôi định nghĩa bao gồm đầu vào là toàn bộ thông tin của một tin tuyển dụng trực tuyến đã đề cập bên trên và ba đầu ra.

**Đầu ra thứ nhất** là nhãn chính cho tin tuyển dụng này bao gồm CLEAN, SEEDING hoặc WARNING. Đối với từng nhãn sẽ được định nghĩa như sau:

- **CLEAN:** Tin tuyển dụng chứa các thông tin rõ ràng, gần như đầy đủ thông tin cần thiết và không chứa các thông tin nhiễu, gây nhầm lẫn.
- **SEEDING:** Tin tuyển dụng chứa các thông tin bổ sung gây nhầm lẫn, không cần thiết với nhiều mục đích khác như thu hút người xem, quảng bá, vv. Các tin tuyển dụng này có thể không phải là giả hay lừa đảo nhưng dễ gây hiểu lầm về quyền lợi, lợi ích, vv của người lao động.
- **WARNING:** Tin tuyển dụng không rõ ràng, không minh bạch, có bất thường hoặc có thông tin bị sai lệch, bị thiếu nhất quán. Đây là các tin tuyển dụng này có khả năng cao là các tin tuyển dụng giả hoặc lừa đảo mà người lao động cần cảnh giác.

**Đầu ra thứ hai** là mức độ xác thực được đánh giá đối với từng khía cạnh. Mức độ đánh giá này có thể bị phụ thuộc bởi thông tin của các khía cạnh khác. Tương tự với bài toán Aspect Category-based Sentiment Analysis, bài toán sẽ bao gồm 5 khía cạnh với các mức độ như sau:

- **POSITIVE:** Chứa đầy đủ các thông tin quan trọng, nhất quán và được xác thực.
- **NEGATIVE:** Ngược lại với positive, có chứa các thông tin gây nhầm lẫn, không rõ ràng, không minh bạch hoặc không xác thực.
- **NOT-MENTIONED:** Thông tin không được cung cấp, hoặc không được đề cập

đến gây cản trở cho việc đánh giá.

- NEUTRAL: Không thuộc ba nhãn trên hoặc không thể kết luận.

**Đầu ra thứ ba** của bài toán là một đoạn văn bản giải thích được viết bằng ngôn ngữ tự nhiên tiếng Việt, nhằm giải thích tại sao mẫu dữ liệu thuộc nhãn CLEAN, SEEDING hoặc WARNING một cách chi tiết.

### Nội dung 3: Phương pháp tiếp cận

Chúng tôi đề xuất phương pháp sử dụng các mô hình pre-trained BERT-based transformers trên ngôn ngữ tiếng Việt để trích xuất các đặc trưng ngữ nghĩa, sau đó đưa các vector ngữ nghĩa qua lớp fully connected (FCs) và output là dự đoán. Cụ thể, chúng tôi sử dụng ViSoBERT [5], PhoBERT, XLM-R, DistilBERT, CafeBERT.

Đầu vào của mô hình được thiết kế ở dạng chuỗi văn bản, với các thông tin của tin tuyển dụng được nối với nhau, phân cách bởi các token khác nhau đối với từng mô hình pre-trained. Sau khi được tokenize thành các id của chuỗi văn bản đầu vào, chuỗi id được padding theo độ dài lớn nhất có thể xử lý được của từng mô hình pre-trained, bằng các padding token riêng biệt. Cấu trúc dữ liệu đầu vào tương tự sau đây:

*< cls > feature\_1 < sep > feature\_2 < sep > ... feature\_n ... < sep > < padding > ...*

Đối với tác vụ phân lớp cho đầu ra thứ nhất, sau khi sử dụng mô hình pre-trained để embedded dữ liệu đầu vào, thì các đặc trưng này sẽ được đưa vào các lớp fully connected (FCs). Đồng thời, chúng tôi sử dụng thêm các lớp RELU và Dropout để giảm thiểu overfitting. Ở lớp fully connected cuối cùng, lớp Softmax được sử dụng để tính xác suất của từng nhãn dự đoán. Đối với tác vụ phân tích mức độ xác thực thông tin của đầu ra thứ hai, chúng tôi sử dụng phương pháp và kiến trúc gần giống với tác vụ trước đó. Tuy nhiên, ở lớp fully connected cuối, chúng tôi sử dụng một module list các FCs bằng số lượng khía cạnh thay vì các node đơn lẻ. Cuối cùng, đối với tác vụ tạo sinh câu giải thích ở đầu ra cuối cùng, chúng tôi sử dụng các mô hình tạo sinh ngôn ngữ như BARTpho [6], ViT5.

## KẾT QUẢ MONG ĐỢI

Tổng kết, sau đây là các kết quả mong đợi của nhóm nghiên cứu:

- Xây dựng guideline gán nhãn cho các mẫu tin tuyển dụng lừa đảo, cần cảnh giác có thể tham khảo.
- Xây dựng bộ dữ liệu đánh giá khía cạnh tin tuyển dụng với tối thiểu 10.000 mẫu tin.
- Xây dựng các mô hình baseline (gồm hướng tiếp cận đã nêu và thử nghiệm thêm các phương pháp XAI mới như LIME, SHAP, ANCHOR) trên tập dữ liệu đề xuất.

## **TÀI LIỆU THAM KHẢO** (*Định dạng DBLP*)

1. Stefanos Vidros, Constantinos Kolias, Georgios Kambourakis: Online recruitment services: another playground for fraudsters. *Comput. Fraud Secur.* 2016(3): 8-13 (2016)
2. Stefanos Vidros, Constantinos Kolias, Georgios Kambourakis, Leman Akoglu: Automatic detection of online recruitment frauds: Characteristics, methods, and a public dataset. *Future Internet* 9(1): 6 (2017)
3. Dang Van Thin et al.: Multi-task learning for aspect and polarity recognition on Vietnamese datasets. *PACLIC* 2019
4. Tin Van Huynh, Kiet Van Nguyen, Ngan Luu-Thuy Nguyen, Anh Gia-Tuan Nguyen: Job prediction: From deep neural network models to applications. *RIVF* 2020: 1-6
5. Quoc-Nam Nguyen, Thang Chau Phan, Duc-Vu Nguyen, Kiet Van Nguyen: ViSoBERT: A Pre-Trained Language Model for Vietnamese Social Media Text Processing. *CoRR* abs/2310.11166 (2023)
6. Nguyen Luong Tran, Duong Minh Le, Dat Quoc Nguyen: BartPho: pre-trained sequence-to-sequence models for Vietnamese. *CoRR* abs/2109.09701 (2021)