# **THÔNG TIN CHUNG CỦA BÁO CÁO**

* Link YouTube video của báo cáo (tối đa 5 phút):   
  [*https://www.youtube.com/watch?v=IZJR5g5RxRg*](https://www.youtube.com/watch?v=IZJR5g5RxRg)
* Link slides (dạng .pdf đặt trên Github):   
  [*https://github.com/PhuongNT5/CS2205.MAR2024/blob/main/Su%20dung%20hoc%20tap%20lien%20ket%20co%20giam%20sat%20de%20tang%20trai%20nghiem%20nguoi%20dung%20mang%20xa%20hoi.pdf*](https://github.com/PhuongNT5/CS2205.MAR2024/blob/main/Su%20dung%20hoc%20tap%20lien%20ket%20co%20giam%20sat%20de%20tang%20trai%20nghiem%20nguoi%20dung%20mang%20xa%20hoi.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*

| * Họ và Tên: Nguyễn Thị Phương * MSSV: 230201051 | * Lớp: CS2205.CH181 * Tự đánh giá (điểm tổng kết môn): * Số buổi vắng: 1 * Số câu hỏi QT cá nhân: * Link Github: |
| --- | --- |

# **ĐỀ CƯƠNG NGHIÊN CỨU**

| **TÊN ĐỀ TÀI (IN HOA)**  FEDERATED SELF-SUPERVISED LEARNING TO IMPROVE SOCIAL MEDIA USER EXPERIENCE |
| --- |
| **TÊN ĐỀ TÀI TIẾNG ANH (IN HOA)**  SỬ DỤNG HỌC TẬP LIÊN KẾT GIÁM SÁT ĐỂ TĂNG TRẢI NGHIỆM NGƯỜI DÙNG MẠNG XÃ HỘI |
| **TÓM TẮT** *(Tối đa 400 từ)*  Mỗi ngày, hàng triệu nội dung video được tải lên mạng xã hội. Tuy nhiên tình trạng và chất lượng video đa dạng đôi khi gặp phải các vấn đề quản lý nội dung và đảm bảo tính an toàn cho người dùng. Một số vấn đề thường gặp bao gồm nội dung không phù hợp, bạo lực, khiêu dâm, vi phạm bản quyền và tin tức giả mạo. Các hành động không phù hợp trong video có thể gây hại đến người xem và gây ảnh hưởng tiêu cực đến trải nghiệm người dùng mạng xã hội. Vì thế trong đề tài này, phương pháp Federated Self-Supervised Learning (FedVSSL) được đề xuất để giải quyết vấn đề đó. Phương pháp tự học liên kết có giám sát là một phương pháp học máy, trong đó mô hình được huấn luyện bằng cách sử dụng dữ liệu không gắn nhãn từ nhiều nguồn khác nhau. Bằng cách sử dụng phương pháp này với dữ liệu mạng xã hội, nghiên cứu nhằm tạo ra một hệ thống tự động gợi ý, hạn chế nội dung, cải thiện trải nghiệm người dùng trên mạng xã hội. Không chỉ thế, dữ liệu người dùng sẽ không cần phải chia sẻ trực tiếp mà chỉ sử dụng để cập nhật mô hình phân tích trên các thiết bị phân tán. Điều này giúp bảo vệ thông tin cá nhân và đáp ứng các yêu cầu về quyền riêng tư. |
| **GIỚI THIỆU** *(Tối đa 1 trang A4)*  Mạng xã hội đã trở thành một phần không thể thiếu trong cuộc sống hàng ngày của chúng ta, đóng vai trò quan trọng trong việc chia sẻ thông tin, giao tiếp và giải trí. Trên mạng xã hội, video là một yếu tố quan trọng để truyền tải nội dung và tạo trải nghiệm đa dạng cho người dùng.  Điều này đặt ra một thách thức đối với các nhà quản lý nội dung và các chính phủ trên khắp thế giới, khi họ phải tìm cách hiệu quả để giám sát và kiểm soát nội dung trên các nền tảng truyền thông xã hội.  Đối phó với các thách thức về vấn đề video này đã được các nhà nghiên cứu quan tâm từ lâu trong thị giác máy tính. Học tự giám sát (Self-supervised Learning) là phương pháp trong học máy cho phép mô hình học từ dữ liệu mà không yêu cầu nhãn được cung cấp trước, nhãn này được tạo ra từ dữ liệu đầu vào. Nó là nền tảng để giải quyết các vấn đề nhận diện hành động của con người, phát hiện thời gian của hành động,... Tuy nhiên, trên thực tế, việc triển khai SSL đơn thuần đòi hỏi một lượng lớn dữ liệu tập trung ở máy chủ để xử lý. Điều này gây ra những lo ngại đáng kể về quyền riêng tư, chi phí liên lạc và lưu trữ, giới hạn ở các tập dữ liệu nhỏ.  Trong đề tài này, chúng tôi đề xuất sử dụng mô hình Học tập liên kết có giám sát để phân tích video – Federated Self-supervised Learning for Video Understanding (FedVSSL). Nó kết hợp video – SSL với FL đào tạo mang lại hiệu suất tốt hơn nhiều so với SSL thuần trong nhiệm vụ truy xuất dữ liệu và nhận dạng hành động. Mô hình này không chỉ giúp tăng cường khả năng tùy biến trên từng máy khách mà còn mang lại tính bảo mật và hiệu suất vượt trội so với các phương pháp truyền thống khác. FedVSSL hứa hẹn là một công cụ mạnh mẽ trong việc nhận diện và loại bỏ nội dung độc hại trên các nền tảng truyền thông xã hội.  Phương pháp học tập liên kết giám sát kết hợp việc sử dụng dữ liệu từ nhiều nguồn khác nhau và mô hình học máy để xây dựng một hệ thống thông minh có khả năng nhận dạng và phân tích hành động trong video.  Bằng cách áp dụng phương pháp này, báo cáo đề xuất một quy trình phân tích hành động video hiệu quả và tự động trên mạng xã hội. Quá trình này bao gồm thu thập và chuẩn bị dữ liệu, xây dựng và huấn luyện mô hình học tập liên kết giám sát và áp dụng mô hình để phân tích và nhận dạng hành động trong video trực tuyến. |
| **MỤC TIÊU**   * Tìm hiểu và nghiên cứu mô hình học tập liên kết có giám sát FedVSSL (Federated Self-supervised Learning). Phân tích trên các khía cạnh về cải thiện hiệu suất, độ đảm bảo an toàn trên từng máy khách mà 1 mô hình Federated Learning mang lại. * Huấn luyện và đánh giá mô hình FedVSSL: Sử dụng bộ dữ liệu đã thu thập để huấn luyện lại mô hình FedVSSL. Ghi nhận và đánh giá đây là một phương pháp mạnh mẽ để phân tích và nhận diện video trên mạng xã hội đảm bảo an toàn. * Xây dựng một mô hình học máy phân tán và bảo mật: Mục tiêu quan trọng khác là xây dựng một mô hình học máy phân tán và bảo mật trong quá trình huấn luyện và triển khai thuật toán FedVSSL. Điều này đảm bảo sự bảo mật và quyền riêng tư cho dữ liệu người dùng mạng xã hội * Cải thiện trải nghiệm người dùng trên mạng xã hội: Một mục tiêu quan trọng khác của nghiên cứu là trải nghiệm người dùng trên mạng xã hội thông qua việc phân tích và nhận diện hành động trong video. Từ đó, hệ thống có thể gợi ý nội dung phù hợp, tăng cường tương tác và cung cấp trải nghiệm cá nhân hóa cho người dùng. |
| **NỘI DUNG VÀ PHƯƠNG PHÁP**  *(Viết nội dung và phương pháp thực hiện để đạt được các mục tiêu đã nêu)*   * Tìm hiểu mô hình học tập liên kết (Federated Learning): một mô hình cho phép học máy trên dữ liệu phân tán trên nhiều thiết bị hoặc thực thể hợp tác dưới một mô hình dùng chung mà không cần trao đổi dữ liệu với nhau.   Cách tiếp cận này cho phép học máy bảo đảm quyền riêng tư vì mỗi thiết bị lưu trữ dữ liệu cục bộ và chỉ chia sẻ thông tin cần thiết đến để cải thiện mô hình.  A diagram of a computer process  Description automatically generated   * Tìm hiểu mô hình học tập có giám sát (Self-supervised Learning): tận dụng phương pháp để học và rút trích đặc trưng từ dữ liệu video. Học tập tự giám sát cho phép mô hình học từ dữ liệu không có nhãn thông qua việc đặt các nhiệm vụ phụ, chẳng hạn như dự đoán khung tiếp theo trong một chuỗi video. * Thu thập dữ liệu video: tìm kiếm các bộ dữ liệu video có sẵn trên mạng xã hội để thu thập các video đa dạng và phong phú. Tạo ra các nhiệm vụ học tự giám sát từ dữ liệu video. * Tiền xử lý dữ liệu: dữ liệu video thu thập được sẽ được tiền xử lý để chuẩn hóa định dạng, giảm nhiễu và chuẩn bị cho quá trình huấn luyện mô hình. * Huấn luyện mô hình: tiến hành quá trình huấn luyện mô hình sử dụng dữ liệu video phân tán và nhiệm vụ học tự giám sát đã xác định. Quá trình này nhằm cải thiện khả năng nhận diện và hiểu video của mô hình. * Học tập liên kết có giám sát: sau khi mô hình được huấn luyện nó sẽ được cập nhật thông qua quá trình học tập có giám sát. Các thiết bị phân tán trên mạng xã hội sẽ gửi các đặc trưng đã học từ dữ liệu video của người dùng tới một trung tâm tập trung. Trung tâm sẽ sử dụng các đặc trưng này để cập nhật mô hình phân tích hành động và nhận diện video. e. Đánh giá và tối ưu hóa: Mô hình được đánh giá thông qua các bài kiểm tra và so sánh với các phương pháp khác trong việc phân tích hành động và nhận diện video trên mạng xã hội. Các tham số và siêu tham số của mô hình cũng được điều chỉnh để tối ưu hóa hiệu suất và độ chính xác. * Áp dụng trên mạng xã hội: tiến hành triển khai trên mạng xã hội để phân tích hành động và nhận diện video. Kết quát có thể sử dụng để cung cấp trải nghiệm người dùng cá nhân hóa, gợi ý nội dung phù hợp và tăng tương tác trên mạng xã hội. |
| **KẾT QUẢ MONG ĐỢI**   * Hiệu quả và độ chính xác cao trong việc nhận diện và phân tích hành động trong video: Dự kiến thuật toán FedVSSL sẽ có khả năng tự động học và rút trích đặc trưng từ dữ liệu video, từ đó phát hiện các video có hành động mang tích chất không lành mạnh như bạo lực, khiêu dâm, vi phạm bản quyền để cảnh báo hoặc ngăn chặn khỏi hệ thống mạng xã hội. * Quá trình phân loại và nhận diện video tự động và hiệu quả: điều này giúp tối ưu hóa thời gian đối với việc phân tích video trên mạng xã hội, giảm thiểu sự can thiệp của con người. * Bảo mật và quyền riêng tư của dữ liệu người dùng được đảm bảo: Sử dụng phương pháp học tập liên kết, người dùng sẽ không cần phải chia sẻ trực tiếp dữ liệu cá nhân của mình, mà chỉ được sử dụng để cập nhật mô hình phân tích trên các thiết bị phân tán. Điều này giúp bảo vệ thông tin cá nhân và đáp ứng các yêu cầu về quyền riêng tư trên mạng xã hội. |
| **TÀI LIỆU THAM KHẢO** *(Định dạng DBLP)*  [1] Rehman, Y. A. U., Gao, Y., Shen, J., de Gusmao, P. P. B., & Lane, N. (2022, October). Federated self-supervised learning for video understanding. In *European Conference on Computer Vision* (pp. 506-522). Cham: Springer Nature Switzerland.  [2] Wang, J., Jiao, J., & Liu, Y. H. (2020). Self-supervised video representation learning by pace prediction. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XVII 16* (pp. 504-521). Springer International Publishing.  [3] Kuehne, H., Jhuang, H., Garrote, E., Poggio, T., Serre, T.: HMDB: a large video database for human motion recognition. In Proceedings of the International Conference on Computer Vision (ICCV) (2011)[4] Schiappa, M. C., Rawat, Y. S., & Shah, M. (2023). Self-supervised learning for videos: A survey. *ACM Computing Surveys*, *55*(13s), 1-37.  [5] Han, T., Xie, W., & Zisserman, A. (2020). Self-supervised co-training for video representation learning. In *Advances in neural information processing systems*, *33*, 5679-5690.  [6] Syed Hammad Ahmed, Muhammad Junaid Khan, H. M. Umer Qaisar, Gita Sukthankar (2023). Malicious or Benign? Towards Effective Content Moderation for Children's Videos. In *The International FLAIRS Conference Proceedings. 36, 1*  [7] Gkolemi, M.; Papadopoulos, P.; Markatos, E.; and Kourtellis, N. 2022. YouTubers Not MadeForKids: Detecting channels sharing inappropriate videos targeting children. In 14th ACM Web Science Conference 2022, WebSci ’22, 370–381. New York, NY, USA: Association for Computing Machinery.  [8] Papadamou, K.; Papasavva, A.; Zannettou, S.; Blackburn, J.; Kourtellis, N.; Leontiadis, I.; Stringhini, G.; and Sirivianos, M. 2021. Disturbed YouTube for Kids: Characterizing and detecting inappropriate videos targeting young children. (arXiv:1901.07046). arXiv:1901.07046 [cs]. |