# **THÔNG TIN CHUNG CỦA NHÓM**

* Link YouTube video của báo cáo (tối đa 5 phút): <https://youtu.be/oWWke4GFXAc>
* Link slides (dạng .pdf đặt trên Github của nhóm):   
  <https://github.com/Haielec/CS519.P11.git>
* *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: Vũ Duy Hải * MSSV: 21520812   OwnPic | * Lớp: CS519.P11 * Tự đánh giá (điểm tổng kết môn): 8/10 * Số buổi vắng: 2 * Số câu hỏi QT cá nhân: 2 * Số câu hỏi QT của cả nhóm: 2 * Mô tả công việc và đóng góp của cá nhân cho kết quả của nhóm:   + Tìm hiểu và triển khai đề tài   + Viết đề cương nghiên cứu   + Thiết kế slide   + Thiết kế poster   + Thuyết trình đề tài và làm video |

# **ĐỀ CƯƠNG NGHIÊN CỨU**

|  |
| --- |
| **TÊN ĐỀ TÀI (IN HOA)**  NGHIÊN CỨU VÀ CẢI THIỆN A.I TRONG VIỆC TẠO VÀ CHỈNH SỬA HÌNH ẢNH THEO YÊU CẦU |
| **TÊN ĐỀ TÀI TIẾNG ANH (IN HOA)**  RESEARCH AND IMPROVE A.I IN CREATING AND EDITING IMAGES ON DEMANDS |
| **TÓM TẮT** *(Tối đa 400 từ)*  - Trong đề tài nghiên cứu này, chúng tôi tập trung vào việc nghiên cứu và cải tiến các mô hình trí tuệ nhân tạo (AI) nhằm tăng cường khả năng tạo và chỉnh sửa hình ảnh dựa trên yêu cầu cụ thể của người dùng. Với sự phát triển mạnh mẽ của các mô hình tạo hình ảnh như Generative Adversarial Networks (GAN) và Diffusion Models, nghiên cứu này hướng tới việc nâng cao độ chính xác, tính thẩm mỹ, và mức độ cá nhân hóa trong các hình ảnh được tạo ra. |
| **GIỚI THIỆU** *(Tối đa 1 trang A4)*  - A.I tạo và chỉnh sửa hình ảnh đã là một khái niệm không hề xa lạ. Một người muốn tạo ra hoặc chỉnh sửa một hình ảnh theo ý họ, nhưng họ lại không hề có kinh nghiệm về máy tính, về các thao tác để làm điều đó, trong khi việc thuê người khác tạo và chỉnh sửa giúp đôi khi có kinh phí khá cao so với mục đích đơn giản của họ khi muốn tạo ra hình ảnh, cũng có nhiều lúc khá khó để tìm kiếm và thuê người để làm việc đó cho họ. Vì vậy, việc các phần mềm A.I tạo và chỉnh sửa hình ảnh ra đời là một điều tất yếu.  - Hiện nay, khi tìm kiếm trên các trang trình duyệt, ta dễ dàng bắt gặp những phần mềm tạo và chỉnh sửa ảnh theo thông qua các yêu cầu viết tay hoặc giọng nói; các phần mềm này có cả online, hoặc phải tải xuống, miễn phí, hoặc trả phí. Điểm chung cho các A.I này trong hiện tại là chúng khá thông minh, có thể nhận diện gần như đầy đủ yêu cầu của người dùng, nhưng hình ảnh chúng tạo ra và chỉnh sửa nhiều khi chưa đáp ứng đúng được các yêu cầu, đặc biệt là các yêu cầu khó tính.  - Nhằm giúp cải thiện các vấn đề trên, chúng tôi cố gắng nghiên cứu và đưa ra các giải pháp cụ thể nhất, bao gồm:   * **Tối ưu hóa chất lượng hình ảnh** * **Chỉnh sửa hình ảnh thông minh** * **Hiểu và xử lý yêu cầu của người dùng** |
| **MỤC TIÊU** *(Viết trong vòng 3 mục tiêu)*  **- Tối ưu hóa chất lượng hình ảnh**: Đề xuất các thuật toán mới hoặc cải tiến các mô hình hiện có để tạo ra hình ảnh có độ phân giải cao, chi tiết rõ ràng và đáp ứng yêu cầu phức tạp của người dùng.  - **Chỉnh sửa hình ảnh thông minh**: Tích hợp các công cụ chỉnh sửa hình ảnh tự động, cho phép thay đổi màu sắc, ánh sáng, phong cách hoặc thêm/xóa đối tượng dựa trên mô tả văn bản.  **- Hiểu và xử lý yêu cầu của người dùng**: Tăng cường khả năng hiểu ngữ cảnh và ý định của người dùng thông qua mô hình xử lý ngôn ngữ tự nhiên (NLP). |
| **NỘI DUNG VÀ PHƯƠNG PHÁP**  - **Nâng cao chất lượng hình ảnh tạo ra**:   * Chúng tôi nhận thấy mô hình GAN được sử dụng rất nhiều trong các web, app tạo hình ảnh hiện tại, mô hình này khá tốt, nhưng có nhược điểm là mất cân bằng giữa generator (bộ tạo) và discriminator (bộ phân biệt), dẫn đến hiện tượng "mode collapse" (mô hình chỉ tạo ra một số dạng hình ảnh lặp đi lặp lại). * Trong khi đó, Diffusion Models khắc phục được tình trạng này, đồng thời cũng có khả năng tạo ra hình ảnh với độ phân giải cao hơn, chúng tôi dự định thay thế mô hình GAN bằng Diffusion Models. * Khi đó, sẽ phát sinh thêm vấn đề là Diffusion Models có chi phí tính toán khá cao và yêu cầu nhiều tài nguyên huấn luyện. Do đó, chúng tôi sẽ không thực sự dùng mô hình Diffusion Models mà sẽ từ mô hình này: * Sử dụng các kỹ thuật nén như **LoRA (Low-Rank Adaptation)** để giảm kích thước và tăng tốc mô hình. * Áp dụng **Distillation** (chưng cất tri thức) để chuyển mô hình lớn thành một phiên bản nhỏ hơn mà vẫn duy trì chất lượng.   - **Tăng cường khả năng cá nhân hóa:**   * Nghiên cứu và kết hợp các mô hình xử lý ngôn ngữ tự nhiên (NLP) với thị giác máy tính (CV) để hiểu rõ hơn ý định của người dùng. * Hiện nay, phương pháp trên đã được các web, app áp dụng nhưng vẫn chưa quá hiệu quả, nên việc nghiên cứu thêm và áp dụng thêm các mô hình kết hợp từ các mô hình khác nhau của 2 loại mô hình trên, cải thiện tham số mô hình cho phù hợp, so sánh hiệu suất giữa chúng để chọn ra mô hình tốt nhất là điều cần thiết (cơ sở để cải thiện sẽ là mức độ hài lòng của khách hàng về hình ảnh được tạo ra hoặc chỉnh sửa, mức độ hài lòng của chính chúng tôi trong khi thực nghiệm trên mô hình đó).   - **Cải thiện khả năng chỉnh sửa hình ảnh thông minh:**   * Sử dụng mạng nơ-ron đối nghịch có điều khiển (ControlNet) để cho phép tinh chỉnh các đặc điểm cụ thể trong hình ảnh. * Các mô hình tạo hình ảnh đề cập phía trên cũng có tích hợp thêm các chức năng chỉnh sửa hình ảnh, nhưng vẫn còn khá hạn chế. Chức năng chỉnh sửa hiện tại trên nhiều app, web chỉnh sửa còn sai, chưa phù hợp bối cảnh hình ảnh, ta sẽ cải thiện phần mô hình thị giác máy tính (CV) để phân biệt được rõ ràng hơn các thành phần, vật thể,… có trên hình ảnh đó (không nhất thiết phải là ảnh thực, có thể là ảnh theo phong cách gothic, anime,…), giúp thuận tiện cho việc chỉnh sửa, thêm, bớt vật thể, ánh sáng,… cho từng ảnh theo yêu cầu. |
| **KẾT QUẢ MONG ĐỢI**  - Mô hình sau khi hoàn thiện hoạt động ổn định, thời gian tính toán cho kết quả nhanh, hiệu suất tốt, hiệu quả cao.  - Mô hình đáp ứng được nhu cầu của người dùng trong đa số trường hợp (mức độ hài lòng khoảng 80% trở lên).  - Mô hình sẽ được áp dụng rộng rãi, được tiếp tục đem ra phân tích nghiên cứu để cải thiện hơn. |
| **TÀI LIỆU THAM KHẢO**   1. **Amazon Web Services (AWS)**. "What is Artificial Intelligence?" Amazon Web Services, 2023. [Online]. Available: [https://aws.amazon.com/vi/what-is/artificial-intelligence](https://aws.amazon.com/vi/what-is/artificial-intelligence" \t "_new). 2. **Intel**. "What is Generative AI?" Intel Corporation, 2023. [Online]. Available: <https://www.intel.vn/content/www/vn/vi/learn/what-is-generative-ai.html.> 3. **Liên hiệp các Hội Khoa học và Kỹ thuật Thừa Thiên Huế (HUSTA)**. "Trí tuệ nhân tạo (AI) và những ứng dụng thực tế của AI trong đời sống." HUSTA, 2024. [Online]. Available: <https://husta.vn/tri-tue-nhan-tao-ai-va-nhung-ung-dung-thuc-te-cua-ai-trong-doi-song.> 4. **Digi-Texx**. "Công Nghệ AI Đang Cách Mạng Hóa Dịch Vụ Xử Lý Ảnh Như Thế Nào?" Digi-Texx, 2024. [Online]. Available: <https://digi-texx.com/vi/techblog-vi/ai-dang-cach-mang-hoa-dich-vu-xu-ly-anh-nhu-the-nao.> 5. **The Atlantic**. "Why Does AI Art Look Like That?" The Atlantic, 2024. [Online]. Available: <https://www.theatlantic.com/technology/archive/2024/08/why-does-all-ai-art-look-same/679488/.> |