

ĐẠI HỌC QUỐC GIA TP.HCM
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN



KHOA KHOA HỌC MÁY TÍNH

MÔN HỌC: CÁC KĨ THUẬT HỌC SÂU VÀ ỨNG DỤNG - CS431.Q11

BÁO CÁO ĐỒ ÁN - LẬP NHÓM

Sinh viên thực hiện:

1. Nguyễn Chí Cường – 23520199
2. Lê Thành Thắng Dạt – 23520251
3. Vũ Gia Khang – 23520713
4. Nguyễn Đăng Khoa – 23520746

Giảng viên:
TS. Nguyễn Vinh Tiệp

Contents

1	Danh sách thành viên nhóm	2
2	Đề tài	2
3	Phân công công việc	2
4	Kết quả dự kiến đạt được	3

1 Danh sách thành viên nhóm

Họ và tên	MSSV	Vai trò
Nguyễn Chí Cường	23520199	Trưởng nhóm
Lê Thành Thắng Dạt	23520251	Thành viên
Vũ Gia Khang ^(*)	23520713	Thành viên
Nguyễn Đăng Khoa	23520746	Thành viên

(*) Bạn Vũ Gia Khang đã email cho giảng viên về vấn đề làm nhóm 4 người.

2 Đề tài

Trong bối cảnh công nghệ trí tuệ nhân tạo (AI) phát triển mạnh mẽ, các mô hình tạo sinh (Generative Models) như **Stable Diffusion** hay **Midjourney** đã mang lại khả năng tạo ra hình ảnh chất lượng cao chỉ từ văn bản mô tả. Nhưng những mô hình AI này thường **khó giữ nguyên đặc trưng khuôn mặt người dùng** và **mất nhiều thời gian huấn luyện**. InstantID cho phép tạo ảnh cá nhân hóa **nhanh, chính xác và không cần fine-tuning**, giúp **tối ưu trải nghiệm người dùng** trong các ứng dụng như tạo avatar, nghệ thuật số hay nền tảng mạng xã hội. Vì vậy, nhóm chọn đề tài này để nghiên cứu và ứng dụng công nghệ **tạo sinh giữ danh tính trong thời gian thực**.

3 Phân công công việc

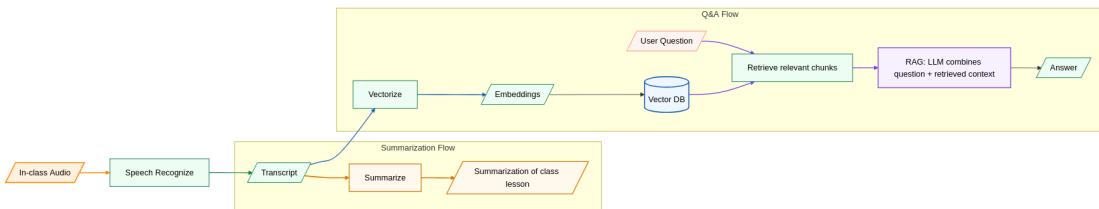


Figure 1: Pipeline dự kiến của hệ thống

Các node quan trọng trong pipeline gồm:

- ID Embedding (Face Embedding): Ảnh khuôn mặt đầu vào được đưa qua một mô hình nhận diện khuôn mặt (ví dụ antelopev2 hoặc InsightFace) để trích xuất vector đặc trưng (embedding). Vector này chứa thông tin nhận dạng duy nhất của người đó.
- Image Adapter (Lightweight Adapted Module): Một mạng nồng (lightweight adapter) chuyển embedding khuôn mặt thành tín hiệu hướng dẫn (visual prompt) cho mô hình khuếch tán. Mục tiêu là giúp mô hình “hiểu” đặc trưng khuôn mặt mà không cần fine-tuning.
- IdentityNet (ID controller): Module chuyên biệt dùng để đảm bảo toàn danh tính. Nó kết hợp embedding khuôn mặt và tín hiệu từ prompt để điều khiển các tầng attention của mô hình khuếch tán, giữ ổn định đặc trưng khuôn mặt trong suốt quá trình sinh ảnh.
- Base model: Dựa trên mô hình SD1.5 hoặc SDXL, thực hiện quá trình khuếch tán ngược (denoising) để tạo ra hình ảnh hoàn chỉnh, vừa đúng với mô tả văn bản, vừa giữ nguyên danh tính người trong ảnh gốc.

Bảng phân công công việc:

Họ và tên	MSSV	Công việc
Nguyễn Chí Cường	23520199	Face Encoder (ID Embedding)
Lê Thành Thắng Đạt	23520251	Image Adapter (Visual Prompt)
Vũ Gia Khang	23520713	IdentityNet (ID Controller)
Nguyễn Đăng Khoa	23520746	Base Model (Stable Diffusion) + Kết hợp Pipeline

4 Kết quả dự kiến đạt được

Những kết quả dự kiến nhóm sẽ đạt được bao gồm:

- Một demo về hệ thống của nhóm với ảnh được lấy từ nguồn [thispersondoesnotexist](https://thispersondoesnotexist.com)
 - Metric cho bài toán sinh ảnh:
 - * **Identity Similarity (ID Score):** Cosine similarity giữa embedding khuôn mặt gốc và khuôn mặt trong ảnh sinh ra (đo bằng ArcFace).

- * **Image Quality:** FID (Fréchet Inception Distance) hoặc CLIP Score để đánh giá độ chân thực và sự ăn khớp với prompt..
- Một báo cáo chi tiết về quá trình thực hiện dự án, so sánh các mô hình đã sử dụng, các thử nghiệm và đánh giá mô hình.