

Các Mục Đã Được Thay Đổi Và Bổ Sung Thêm

Nhóm: 11

Tên đồ án: AI photofun Studio

Đề tài: Ứng dụng sinh ảnh nhiều mục đích

1. Tổng quan

Trong quá trình phát triển nhóm em cảm thấy số feature vẫn chưa phong phú và gặp một vài vấn đề trong luồng logic đã đưa vì thế đã có một số thay đổi và các feature bổ sung thêm

2. Các chức năng thay đổi

- Mô tả

Chức năng cho phép người dùng nhập mô tả văn bản (prompt) để AI tạo ra hình ảnh tương ứng.

- Lý do thay đổi

Prompt của người dùng thường rất ngắn và thiếu chi tiết (ví dụ: "con mèo"), dẫn đến ảnh tạo ra không đẹp. Ngoài ra, cách xử lý đồng bộ (synchronous) cũ khiến người dùng phải chờ đợi lâu và giao diện bị treo.

- Flow

1. Người dùng nhập prompt thô.
2. Hệ thống gửi prompt sang Gemini AI để viết lại chi tiết hơn (thêm ánh sáng, phong cách).
3. Đẩy yêu cầu vào hàng đợi xử lý ngầm (Celery Queue), trả về task_id ngay lập tức.
4. Đẩy yêu cầu vào hàng đợi xử lý ngầm (Celery Queue), trả về task_id ngay lập tức.
5. Lưu kết quả vào Database khi hoàn tất.

- Lợi ích gì

1. Ảnh đầu ra đẹp hơn nhờ prompt được AI tối ưu hóa tự động.
2. Hệ thống không bị nghẽn, người dùng có thể làm việc khác trong khi chờ.

2.2 Conversation Service (Trò chuyện AI)

- Mô tả

Hệ thống chat cho phép người dùng tương tác với AI để điều khiển các tính năng xử lý ảnh.

- **Lý do thay đổi**

Ban đầu chỉ thiết kế như một giao diện chat thông thường để lưu lại lịch sử hội thoại. Nhóm nhận thấy có thể nâng cấp thành AI Agent có khả năng hiểu ý định và tự động thực thi lệnh.

- **Flow**

1. Người dùng nhắm tin (ví dụ: "Hãy upscale bức ảnh này lên 4K").
2. Hệ thống dùng Gemini AI phân tích để xác định ý định (Intent) là "upscale".
3. Tự động trích xuất các tham số cần thiết từ câu nói.
4. Gọi đến Upscale Service để xử lý và trả kết quả ngay trong khung chat.

- **Lợi ích gì**

1. Người dùng chỉ cần ra lệnh bằng ngôn ngữ tự nhiên, không cần tìm kiếm nút bấm.
2. Tích hợp tất cả công cụ vào một chỗ, tiện lợi hơn.

2.3 Tích hợp tất cả công cụ vào một chỗ, tiện lợi hơn.

- **Mô tả**

Chức năng làm nét và tăng độ phân giải ảnh, thay thế cho tính năng "Photo Restore" trong thiết kế ban đầu.

- **Lý do thay đổi**

Khái niệm "Restore" trong thiết kế khá chung chung. Nhóm quyết định tập trung vào kỹ thuật Upscale chuyên sâu với các chế độ xử lý khác nhau cho từng loại ảnh.

- **Flow**

1. Người dùng tải ảnh lên và chọn chế độ (flavor): Photo, Sublime, hoặc Denoiser.
2. Hệ thống áp dụng thuật toán Super-resolution V2 phù hợp để tái tạo chi tiết.
3. Trả về ảnh có độ phân giải cao gấp 2-4 lần.

- **Lợi ích gì**

1. Kết quả sắc nét tự nhiên, tối ưu cho từng loại ảnh cụ thể (ảnh chụp vs tranh vẽ).
2. Tránh tình trạng ảnh bị vỡ hoặc nhìn giả khi phóng to.

2.4 Hệ thống xác thực và quản lý token

- **Mô tả:**

Thay đổi cơ chế xác thực từ single token sang dual token system (Access Token và Refresh Token). Access Token được lưu trữ trong memory, Refresh Token được lưu trữ trong HTTP-only cookie. Tích hợp Axios Interceptors để tự động xử lý authentication và token refresh.

- **Lý do thay đổi:**

Tăng cường bảo mật: Access Token trong memory không thể bị đánh cắp qua XSS attacks, Refresh Token trong HTTP-only cookie không thể bị truy cập bởi JavaScript

Cải thiện trải nghiệm người dùng: Tự động làm mới token mà không cần người dùng đăng nhập lại

Tuân thủ best practices trong việc quản lý authentication tokens

Giảm thiểu rủi ro bảo mật khi token bị lộ (Access Token có thời gian sống ngắn)

- **Flow**

- **Login Flow:**

Người dùng đăng nhập với credentials

Server trả về Access Token (lưu vào memory) và Refresh Token (lưu vào cookie)

Access Token chứa thông tin: token, expiresIn

Refresh Token được set vào cookie với thuộc tính: httpOnly, secure, refreshExpiresIn

- **Request Flow (Axios Request Interceptor):**

Trước mỗi request, interceptor tự động attach Access Token vào header dạng Authorization: Bearer {accessToken}
API yêu cầu authentication sẽ verify Access Token

- **Response Flow (Axios Response Interceptor):**

- a. Nếu response trả về status 401 (Unauthorized): Gọi API refresh token với Refresh Token từ cookie
- b. Nếu Refresh Token còn hiệu lực:
 - i. Server trả về Access Token mới
 - ii. Server thu hồi (revoke) Refresh Token cũ và tạo Refresh Token mới
 - iii. Client lưu Access Token mới vào memory và Refresh Token mới vào cookie
 - iv. Retry request ban đầu với Access Token mới
- c. Nếu Refresh Token hết hiệu lực hoặc không hợp lệ:
 - i. Server trả về 401
 - ii. Client tự động logout và redirect về trang đăng nhập

- **Logout Flow:**

Clear Access Token từ memory

Clear Refresh Token từ cookie

Redirect về trang login

- **Lợi ích:**

- + Bảo mật cao hơn: Token rotation giúp giảm thiểu rủi ro khi token bị compromise. Access Token ngắn hạn hạn chế thời gian khai thác nếu bị lộ
- + Trải nghiệm người dùng mượt mà: Tự động refresh token trong background, người dùng không bị gián đoạn khi làm việc
- + Quản lý session linh hoạt: Có thể thu hồi quyền truy cập ngay lập tức thông qua việc invalidate Refresh Token
- + Giảm tải server: Access Token được verify locally (JWT), chỉ cần gọi server khi refresh
- + Centralized authentication logic: Tất cả logic xác thực được xử lý tập trung tại interceptor, code clean hơn

3. Các chức năng bổ sung

3.1 Video Generation (Tạo Video AI)

- **Mô tả**

Chức năng cho phép tạo video ngắn từ một bức ảnh tĩnh hoặc từ văn bản kịch bản. Đây là tính năng hoàn toàn mới, không có trong thiết kế ban đầu.

- **Flow** như nào

1. Người dùng tải ảnh lên hoặc nhập mô tả kịch bản.
2. Hệ thống gửi yêu cầu sang Model Video.
3. Theo dõi tiến độ xử lý (polling) vì tạo video tốn thời gian.
4. Tải video kết quả về và lưu vào thư viện.

- **Công nghệ**

Model Studio Video Generation API (Veo-2), Celery Long-running Tasks.

3.2 Relight (Chỉnh sửa ánh sáng)

- **Mô tả**

Công cụ AI cho phép thay đổi nguồn sáng và hướng sáng trong ảnh 2D. Ví dụ: đổi ảnh chụp ban ngày thành hoàng hôn.

- **Flow**

1. Người dùng tải ảnh lên và mô tả ánh sáng mong muốn (hoặc chọn ảnh tham chiếu).
2. AI phân tích chiều sâu (Depth Map) của ảnh.
3. Giả lập nguồn sáng mới chiếu vào và tái tạo ảnh.

- **Công nghệ**

Freepik Relight API, Depth-aware Relighting Model.

3.3 Reimagine (Biến đổi ảnh sáng tạo)

- **Mô tả**

Công cụ AI vẽ lại bức ảnh dựa trên mô tả mới nhưng vẫn giữ nguyên cấu trúc và bối cảnh gốc.

- **Flow**

1. Người dùng tải ảnh gốc và nhập mô tả biến đổi (ví dụ: "biến thành tranh anime").
2. Chọn mức độ sáng tạo (Imagination Level): Subtle, Vivid, hoặc Wild.
3. AI sử dụng kỹ thuật Image-to-Image để vẽ lại ảnh theo yêu cầu.

- **Công nghệ**

Freepik Flux Model, Image-to-Image Diffusion.

3.4 Image Expand (Mở rộng khung hình)

- **Mô tả**

Công cụ Outpainting tự động "sáng tác" thêm nội dung để mở rộng biên ảnh theo các hướng mong muốn.

- **Flow**

1. Người dùng tải ảnh và chỉ định hướng mở rộng (trái, phải, trên, dưới) và số pixel.
2. AI phân tích vùng biên của ảnh gốc.
3. Tự động vẽ thêm chi tiết mới phù hợp với ngũ cảnh ảnh.

- **Công nghệ**

Freepik Image Expand API, Generative Outpainting.

3.5 Rec Prompt (Gợi ý Prompt thông minh)

- **Mô tả**

Hệ thống Machine Learning gợi ý các câu mô tả (prompt) khi người dùng đang nhập liệu, dựa trên lịch sử và sở thích cá nhân.

- **Flow**

1. Khi người dùng bắt đầu gõ, hệ thống chuyển đổi văn bản thành vector toán học (TF-IDF)
2. Tính toán độ tương đồng với database prompt mẫu.
3. Lọc kết quả dựa trên "gu" của người dùng (User Profile được cập nhật liên tục).
4. Trả về danh sách gợi ý phù hợp nhất.

- **Công nghệ**

Scikit-learn (TF-IDF Vectorizer), Cosine Similarity, EMA User Profiling.

3.6 Video Call (Cuộc gọi video)

- **Mô tả:** Tính năng gọi video peer-to-peer cho phép người dùng thực hiện cuộc gọi video trực tiếp với nhau. Sử dụng Firebase Realtime Database để signaling và WebRTC API để thiết lập kết nối trực tiếp giữa các browser.

- **Flow:**

a. **Khởi tạo cuộc gọi:**

- i. Người gọi (Caller) tạo RTCPeerConnection và lấy local media stream (camera, microphone) thông qua getUserMedia()
- ii. Caller tạo offer SDP (Session Description Protocol) và set vào localDescription
- iii. Caller lưu offer SDP vào Firebase Realtime Database tại node của cuộc gọi

b. **Signaling qua Firebase:**

- i. Người nhận (Callee) lắng nghe sự kiện trên Firebase và nhận offer SDP

- ii. Callee tạo RTCPeerConnection và set remote description từ offer SDP nhận được
- iii. Callee tạo answer SDP và lưu vào Firebase
- iv. Caller nhận answer SDP từ Firebase và set vào remote description

c. ICE Candidate Exchange:

- i. Mỗi peer thu thập ICE candidates (các đường dẫn kết nối khả dụng) thông qua onicecandidate event
- ii. ICE candidates được gửi qua Firebase Realtime Database
- iii. Mỗi bên nhận và add ICE candidates của đối phương vào RTCPeerConnection

d. Thiết lập kết nối P2P:

- i. WebRTC tự động thương lượng và chọn đường kết nối tối ưu nhất
- ii. Khi kết nối thành công, media streams được truyền trực tiếp peer-to-peer
- iii. Video và audio được render ra các <video> elements

e. Kết thúc cuộc gọi:

- i. Đóng RTCPeerConnection
- ii. Dừng tất cả media tracks
- iii. Xóa dữ liệu signaling trên Firebase
- iv. Reset UI về trạng thái ban đầu

- Công nghệ

- + WebRTC API: Cung cấp khả năng truyền media và data theo thời gian thực giữa các browsers:
- + Firebase Realtime Database:
- + STUN/TURN Servers (Tự thiết lập trên vps)