

Kỹ thuật Prompt trong Thực tế: Cách AI biến đổi 'Sultans of Swing' thành 'The Alley Cats of Swing'

Bởi Frank , 19 Tháng 9 2026

Kỹ thuật Prompt trong Thực tế: Cách AI biến đổi 'Sultans of Swing' thành 'The Alley Cats of Swing'



Bởi Frank, ngày 19 tháng 9 năm 2026

Khả năng đồ họa của trí tuệ nhân tạo hiện nay gần như là vô hạn. Như chúng ta đã thảo luận gần đây trong bài viết về lý do AI nguồn mở và các hệ điều hành mới thay thế các công cụ truyền thống như Photoshop, quyền lực đang chuyển dịch từ các gói phần mềm đắt tiền sang các quy trình làm việc thông minh và dễ tiếp cận. Hôm nay, chúng ta đi sâu hơn vào thế giới của **biến đổi video sang video**. Chúng ta phân tích một nghiên cứu tình huống vui nhộn nhưng ấn tượng về mặt kỹ thuật: sự biến đổi của bản kinh điển *Dire Straits - Sultans Of Swing* thành bản nhại hoạt hình *Dire Strays - The Alley Cats Of Swing*. Làm thế nào để một ban nhạc huyền thoại được hóa thân thành một nhóm mèo đường phố đang chơi nhạc mà không làm mất đi bầu không khí biểu tượng và chuyển động gốc? Câu trả lời nằm ở Kỹ thuật Prompt (Prompt Engineering) tiên tiến.

Sự Biến Đổi: Từ Video 1 sang Video 2

Thách thức trong các dự án như thế này không phải là tạo ra một con mèo ngẫu nhiên, mà là duy trì *tính nhất quán thời gian* (chuyển động xuyên suốt thời gian) và bố cục ban đầu của video. AI phải hiểu rằng người chơi guitar không chỉ là một con mèo bất kỳ, mà là một con mèo thực hiện *chính xác các động tác chơi guitar giống hệt*, dưới cùng ánh sáng và trong cùng bầu không khí câu lạc bộ jazz thập niên 70.

- **Video 1 (Đầu vào):** [Dire Straits - Sultans Of Swing](#)
- **Video 2 (Đầu ra):** [Dire Straits - The Alley Cats Of Swing](#)

Các Công cụ AI Được Sử Dụng Trong Quy Trình

Sự biến đổi như vậy hiếm khi xuất phát từ một 'nút bấm ma thuật' duy nhất. Đó là sự phối hợp của các công cụ chuyên biệt. Dựa trên trạng thái kỹ thuật hiện tại vào năm 2026, quy trình làm việc hiệu quả nhất bao gồm các thành phần sau:

1. Soạn thảo Prompt & Khái niệm hóa

Trước khi các pixel được tạo ra, hướng dẫn phải hoàn hảo. Đây là lúc các mô hình ngôn ngữ với cửa sổ ngữ cảnh lớn phát huy tác dụng, chẳng hạn như [Moonshot AI \(Kimi\)](#). Các mô hình này đặc biệt phù hợp để cấu trúc các mô tả trực quan phức tạp và chi tiết thành các prompt AI khả thi, bao gồm góc máy, ánh sáng và các tham chiếu phong cách.

2. Tạo Ảnh Tham Chiếu (Cơ sở Image-to-Video)

Để xác định diện mạo chính xác của những 'con mèo đường phố', các công cụ tạo ảnh từ văn bản nhanh và lặp lại là không thể thiếu. Các công cụ như [Trình tạo Ảnh AI Perchance](#) cung cấp một môi trường dễ tiếp cận, thường là open-weight, để nhanh chóng kiểm tra xem phong cách nào (ví dụ: 'truyện tranh cổ điển', 'ảnh điện ảnh', hoặc 'lông thú thực tế') hoạt động tốt nhất trước khi chuyển sang các tính toán video nặng nề hơn.

3. Video-sang-Video & Thay Thế Nhân Vật

Phần việc nặng nhọc được thực hiện bởi các mô hình video chuyên biệt. Các tùy chọn đặt tiêu chuẩn trong năm 2026 cho loại chuyển giao phong cách (style transfer) này bao gồm:

- **Stable Video Diffusion (SVD) img2vid:** Một mô hình nguồn mở mạnh mẽ lấy một hình ảnh tham chiếu tĩnh (con mèo đã tạo) và hoạt hình hóa nó dựa trên dữ liệu chuyển động của video gốc.
- **Inpainting Video Tiên Tiến / Hoán đổi Nhân vật:** Các công cụ như Runway (kiến trúc Gen-3) hoặc các giải pháp thay thế nguồn mở với ControlNet, 'khóa' các đường viền và tư thế của các thành viên ban nhạc gốc và chiếu kết cấu mới (con mèo) lên trên đó.

Nghệ thuật của Prompt: Các Ví dụ Cụ Thể

Kỹ thuật Prompt, như chúng ta đã nêu trước đó, là nghệ thuật xây dựng các đầu vào hướng dẫn hệ thống AI đạt được kết quả mong muốn. Một prompt kém sẽ tạo ra một con mèo ngẫu nhiên; một prompt tốt sẽ tạo ra một *nhạc sĩ* vốn dĩ là một con mèo. Dưới đây là các prompt phân lớp cần thiết cho kết quả như vậy:

Prompt Cơ Bản (Phong cách Toàn cầu)

"Cinematic video, 1970s jazz club atmosphere, smoky dim lighting, warm stage lights. Anthropomorphic alley cats playing musical instruments, realistic fur texture, wearing vintage 1970s tailored suits. Maintaining exact original human poses and movements, high temporal consistency, 4k resolution, photorealistic."

Prompt Cảnh Cụ Thể (cho từng thành viên ban nhạc)

Để có kết quả tốt nhất, video thường được xử lý theo từng đoạn hoặc từng nhân vật:

- **Người chơi Guitar:** "Orange tabby cat playing a black electric guitar, standing upright, dynamic fingerpicking motion, vintage suit, focused expression, cinematic rim lighting, shallow depth of field."
- **Người chơi Bass:** "Black and white tuxedo cat playing bass guitar, swaying to the rhythm, dim jazz club background, realistic fur details, professional musician stance."
- **Người chơi Trống:** "Grey tabby cat behind a vintage drum kit, energetic drumming pose, motion blur on drumsticks, dramatic stage lighting, highly detailed fur and whiskers."

'Negative Prompt' (Những gì AI *không* nên làm?)

Việc loại trừ các yếu tố không mong muốn cũng quan trọng không kém: "cartoon, 3d render, anime, deformed paws, extra limbs, morphing, flickering, loss of instrument shape, modern clothing, bright daylight."

Quy Trình Từng Bước cho Nhà Sáng Tạo Độc Lập

Làm thế nào để bạn tự thiết lập điều này? Nền Dân chủ của AI có nghĩa là bạn không cần ngân sách Hollywood, nhưng cần có phương pháp luận:

1. **Trích xuất Khung hình:** Trích xuất các khung hình chính từ video gốc nơi tư thế rõ ràng nhất.
2. **Tạo mẫu:** Sử dụng công cụ như Perchance hoặc Moonshot AI để tạo ra hình ảnh 'mèo mặc vest' hoàn hảo khớp với tư thế trong khung hình.
3. **Biến đổi Video:** Tải video gốc và các hình ảnh tham chiếu đã tạo vào mô hình video-sang-video (như Stable Video Diffusion hoặc một pipeline nguồn mở tương tự). Sử dụng *ControlNet (OpenPose hoặc Depth)* để ép chuyển động của video gốc lên hình ảnh con mèo mới.
4. **Inpainting & Tinh chỉnh:** Sử dụng AI inpainting để sửa lỗi từng khung hình đối với bất kỳ hiện tượng giả tạo nào (ví dụ: bàn tay người ló ra qua lớp lông).
5. **Đồng bộ Âm thanh & Hoàn thiện:** Thêm bản âm thanh gốc, không thay đổi của 'Sultans of Swing' trong một trình chỉnh sửa video truyền thống. Sự hài hước và hiệu ứng nảy sinh chính từ sự tương phản giữa âm nhạc nghiêm túc, quen thuộc và thực tế trực quan phi lý.

Kết luận và Suy ngẫm

Sự biến đổi từ 'Sultans of Swing' sang 'The Alley Cats of Swing' không chỉ là một trò đùa vui nhộn. Nó là minh chứng cho thấy **Kỹ thuật Prompt** và các công cụ AI nguồn mở có thể đảm nhận vai trò của đạo diễn, giám đốc nghệ thuật và chuyên gia hiệu ứng hình ảnh (VFX).

Đối với các nhà báo và nhà sáng tạo độc lập, điều này có nghĩa là sự giải phóng khỏi các hệ sinh thái đóng kín và đắt đỏ. Nó cho phép chúng ta kể những câu chuyện trực quan phức tạp, hoặc trong trường hợp này, thêm một lớp ý nghĩa mới, mang tính phê phán hoặc hài hước vào các di sản văn hóa hiện có. Công nghệ đã có sẵn; thách thức bây giờ nằm ở sự sáng tạo và độ chính xác của các hướng dẫn mà chúng ta đưa ra cho nó. Con mèo đã ra khỏi túi, và nó đang chơi guitar.

<https://youtu.be/IBZgk6q6qEg>