BÁO CÁO DỰ ÁN: ỨNG DỤNG AI TẠO SINH TRONG SÁNG TẠO VIDEO NGẮN VỀ RỦI RO BẢO MẬT

---

1. Giới thiệu

Sự phát triển nhanh chóng của các công cụ AI tạo sinh như ChatGPT, Google Gemini, DALL·E và Canva đã làm thay đổi đáng kể cách con người tiếp cận quá trình sáng tạo nội dung số. Thay vì phải tự xây dựng toàn bộ nội dung từ đầu, người dùng có thể tận dụng AI để tạo bản nháp, hình ảnh minh họa và bố cục thiết kế trong thời gian rất ngắn.

Dự án này lựa chọn xây dựng một video ngắn với chủ đề: "Rủi ro khi sử dụng AI để phát triển website không kiểm duyệt", tập trung vào lỗ hổng SQL Injection. Đây là một trong những lỗ hổng bảo mật phổ biến, có thể cho phép kẻ tấn công truy cập trái phép vào hệ thống, đặc biệt là tài khoản quản trị viên.

Mục tiêu của dự án:

- **Minh họa rõ ràng cả lợi ích và rủi ro khi sử dụng AI trong lập trình**
- **Thực hành sử dụng nhiều công cụ AI khác nhau trong một quy trình hoàn chỉnh**
- **Phân tích vai trò thực sự của con người trong việc kiểm soát chất lượng đầu ra**

Sản phẩm cuối cùng là một video ngắn có tính giáo dục, hướng đến sinh viên công nghệ thông tin, với cấu trúc rõ ràng gồm mở đầu, tình huống tấn công, phân tích và kết luận.

2. Công cụ AI sử dụng

2.1 ChatGPT

- **Vai trò: Viết kịch bản video, tạo nội dung giải thích kỹ thuật**
- **Lý do chọn: Khả năng lập luận tốt, cấu trúc rõ ràng**

2.2 Google Gemini

- **Vai trò: Phản biện nội dung, cải thiện thông điệp**
- **Lý do chọn: Diễn đạt tự nhiên, phù hợp kiểm tra logic nội dung**

2.3 DALL·E

- **Vai trò: Tạo hình ảnh minh họa cho video**
- **Lý do chọn: Tạo ảnh nhanh, phù hợp slide cảnh báo**

2.4 Canva

- **Vai trò: Thiết kế video hoàn chỉnh**
- **Lý do chọn: Dễ sử dụng, tích hợp nhiều thành phần**

3. Quá trình thực hiện

Quy trình được xây dựng theo mô hình:
AI tạo nháp → Con người chỉnh sửa → AI hỗ trợ hình ảnh → Con người kiểm duyệt → AI hỗ trợ thiết kế → Con người hoàn thiện

Tỷ lệ đóng góp:

- **AI: ~55%**
- **Con người: ~45%**

3.1 Sử dụng ChatGPT

Prompt 1:

Viết kịch bản video ngắn (1–2 phút) bằng tiếng Việt về rủi ro khi sử dụng AI để tạo website mà không kiểm tra bảo mật. Nội dung cần bao gồm:

- Mở đầu thu hút

- Tình huống bị tấn công SQL Injection

- Hậu quả (lộ tài khoản admin)

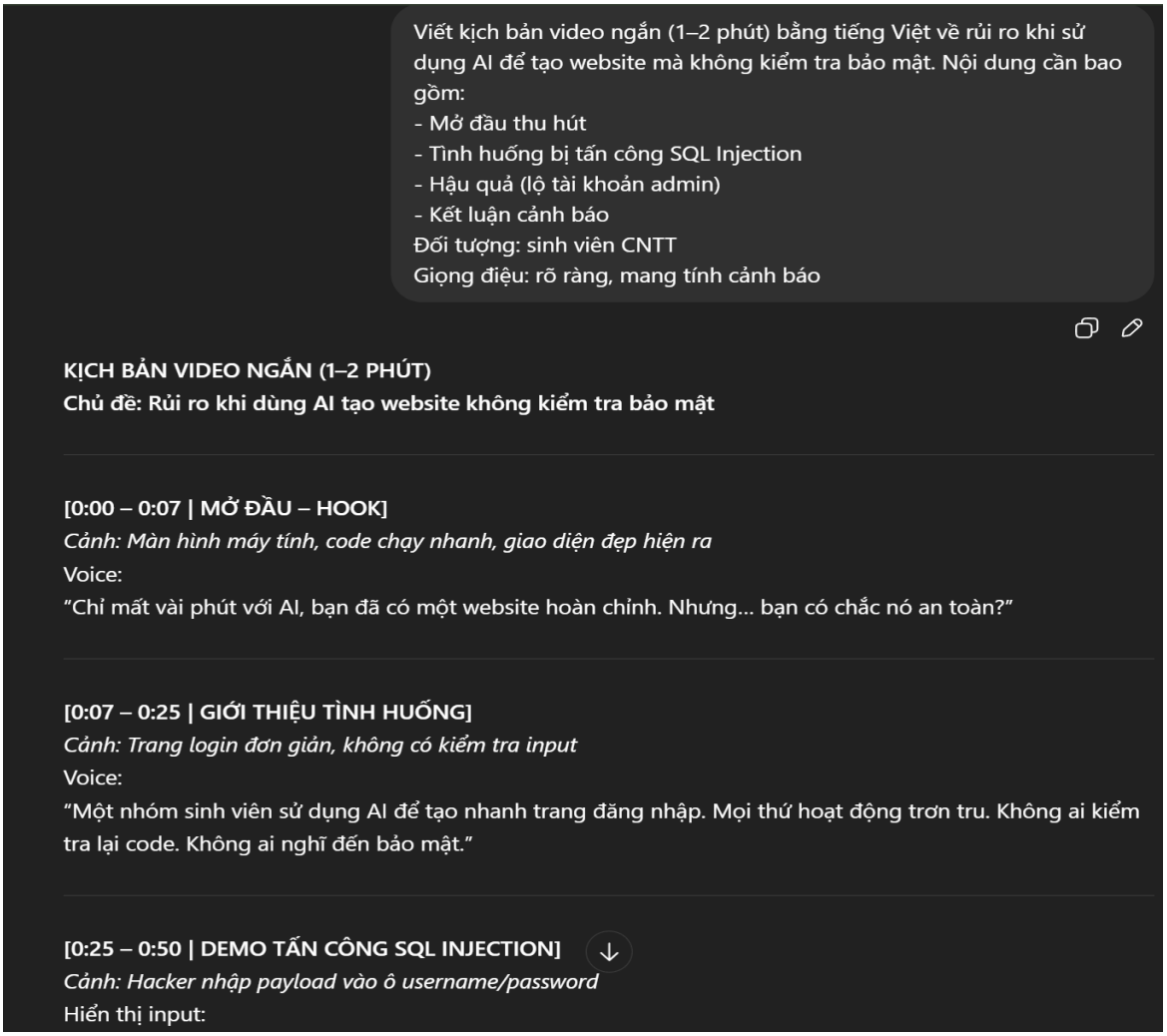
- Kết luận cảnh báo

Đối tượng: sinh viên CNTT

Giọng điệu: rõ ràng, mang tính cảnh báo

Kết quả:

- **AI tạo kịch bản đầy đủ 4 phần**
- **Có ví dụ nhưng còn chung chung**



Cảnh: Nhấn login → truy cập thành công

Voice:
"Và hệ thống... cho phép đăng nhập."

[0:50 – 1:10 | HẬU QUẢ]
Cảnh: Giao diện admin hiện ra, dữ liệu bị truy cập
Voice:
"Kết quả: toàn quyền quản trị bị chiếm. Dữ liệu người dùng bị lộ. Hệ thống bị kiểm soát hoàn toàn."

Cảnh: File database, thông tin user hiện lên

Voice:
"Một lỗi cơ bản. Nhưng hậu quả nghiêm trọng."

[1:10 – 1:25 | PHÂN TÍCH NGẮN]
Cảnh: Highlight đoạn code không dùng prepared statement
Voice:
"AI tạo code nhanh. Nhưng không đảm bảo an toàn. Nếu không hiểu và kiểm tra, bạn đang mở cửa cho tấn công."

[1:25 – 1:40 | KẾT LUẬN – CẢNH BÁO]
Cảnh: Màn hình tối dần, hiện chữ cảnh báo
Voice:
"Đừng tin tưởng mù quáng vào AI. Code có thể chạy. Nhưng bảo mật là trách nhiệm của bạn."

Text trên màn hình:
"AI không chịu trách nhiệm. Bạn thì có."

Prompt 2 (tinh chỉnh):

Viết lại phần tình huống tấn công, bổ sung ví dụ cụ thể về SQL Injection dạng ' OR '1'='1 và giải thích ngắn gọn vì sao nó hoạt động.

Chỉnh sửa của con người:

- **Rút gọn câu để phù hợp video**
- **Thay các đoạn lý thuyết dài bằng ví dụ trực quan**
- **Thêm câu nhấn mạnh trách nhiệm lập trình viên**

Tích hợp:

- **Dùng làm voice-over chính**

- **Dùng làm subtitle**

3.2 Sử dụng Gemini

Prompt 1:

Đánh giá kịch bản video sau về chủ đề AI và bảo mật web. Hãy chỉ ra:

- Điểm chưa rõ

- Điểm chưa thuyết phục

- Cách cải thiện

Kết quả:

- **Gợi ý bổ sung yếu tố "AI có lợi ích"**
- **Nhận xét phần kết luận còn thiên về cảnh báo**

Prompt 2:

Viết lại phần kết luận sao cho cân bằng giữa lợi ích và rủi ro của AI trong lập trình.

Chỉnh sửa:

- **Kết hợp phần viết lại với nội dung gốc**
- **Loại bỏ các câu quá chung chung**

Tích hợp:

- **Dùng cho phần kết luận video**

Đánh giá kịch bản video sau về chủ đề AI và bảo mật web. Hãy chỉ ra:
- Điểm chưa rõ
- Điểm chưa thuyết phục
- Cách cải thiện...

Chào bạn, kịch bản của bạn có cấu trúc rất rõ ràng, nhắm thẳng vào một vấn đề nhức nhối hiện nay: **Sự chủ quan khi dùng AI**. Tuy nhiên, để video thực sự chuyên nghiệp và mang tính xây dựng thay vì chỉ "hù dọa", mình xin đóng góp một vài góc nhìn như sau:

---

1. Những điểm chưa rõ

- **Đối tượng người dùng AI:** Kịch bản chưa làm rõ AI đang được dùng ở đây là gì. Là AI tạo code (như GitHub Copilot, ChatGPT) hay là một nền tảng "AI Website Builder" (no-code)? Cách tiếp cận bảo mật cho hai loại này rất khác nhau.
- **Cơ chế lỗi:** Với khán giả đại chúng, việc chỉ nhìn thấy `' OR '1'='1` có thể hơi khó hiểu. Cần giải thích cực ngắn gọn tại sao đoạn mã đó lại "vượt rào" được (ví dụ: đánh lừa logic so sánh của máy tính).

2. Những điểm chưa thuyết phục

- **Tính thực tế của AI hiện nay:** Các mô hình AI lớn (như GPT-4 hay Claude 3.5) hiện nay thường tự động thêm các biện pháp bảo mật như **Prepared Statements** khi yêu cầu viết code SQL. Việc AI viết ra một đoạn code lỗi SQL Injection cơ bản như thập kỷ trước là có, nhưng đang ít dần. Nếu không giải thích kỹ, người xem có thể thấy ví dụ hơi "cũ".
- **Thiếu sự tương phản:** Video mới chỉ cho thấy mặt tối. Nếu không có sự so sánh giữa "Code AI tạo" và "Code đã qua kiểm soát", người xem sẽ dễ có tâm lý bài trừ AI thay vì học cách dùng đúng.

3. Cách cải thiện & Bổ sung yếu tố "AI có lợi ích"

Để kịch bản cân bằng và thuyết phục hơn, bạn nên lồng ghép việc AI cũng chính là **giải pháp**.

Thay đổi nội dung:

- **Đoạn [1:10 – 1:25]:** Thay vì chỉ nói AI không an toàn, hãy biến AI thành "người gác đền".
  - *Voice mới:* "AI có thể tạo ra lỗi, nhưng chính AI cũng có thể giúp bạn tìm ra lỗi đó. Hãy dùng AI để rà soát lại chính đoạn code nó vừa tạo ra với câu lệnh: 'Kiểm tra lỗ hổng SQL Injection cho đoạn code này'."
- **Bổ sung hình ảnh:** Chèn một cảnh quay AI đưa ra gợi ý sửa lỗi (Fix) sau khi bị tấn công.

Tổng kết đánh giá theo yêu cầu:

1. Gợi ý bổ sung yếu tố "AI có lợi ích"

Đừng chỉ dừng lại ở việc AI tạo ra lỗi. Hãy thêm một bước: **"Dùng AI để kiểm tra bảo mật (Security Review)".**

- *Hành động:* Sau khi thấy lỗi, nhân viên trong video dùng một công cụ AI khác (hoặc chính prompt đó) để yêu cầu: "Optimize this code for security".
- *Thông điệp:* AI là một trợ lý đắc lực, nhưng nó cần một "người sếp" có kiến thức để duyệt kết quả.

2. Nhận xét phần kết luận

Kết luận hiện tại: *"AI không chịu trách nhiệm. Bạn thì có"* mang sắc thái hơi nặng nề và tiêu cực (Fear-mongering).

Cách sửa để mang tính truyền cảm hứng hơn:

- *Text trên màn hình:* **AI + Con người = Bảo mật tối ưu.**
- *Voice kết luận:* "Đừng để tốc độ làm mờ mắt. Hãy dùng AI như một cộng sự, và giữ vai trò kiểm soát cuối cùng. Sáng tạo nhanh, nhưng hãy bảo mật thông minh."

Lời khuyên thêm: Ở đoạn, bạn nên quay cận cảnh con trỏ chuột nhấn vào nút "Login" kèm một hiệu ứng âm thanh "click" đanh gọn để tăng kịch tính trước khi dữ liệu bị lộ. Video ngắn (Shorts/TikTok) rất cần những điểm nhấn âm thanh như vậy!

3.3 Sử dụng DALL·E

Prompt 1:

Tạo hình ảnh minh họa cho video về nguy cơ bảo mật khi dùng AI tạo website. Nội dung gồm:

- Màn hình đăng nhập bị khai thác

- Biểu tượng cảnh báo

- Tông màu tối, phong cách công nghệ

Prompt 2:

Tạo ảnh đơn giản hơn, phù hợp slide thuyết trình, dễ đọc chữ tiếng Việt, bố cục rõ ràng

Kết quả:

- **Ảnh 1: đẹp nhưng phức tạp**
- **Ảnh 2: phù hợp hơn cho slide**

Chỉnh sửa:

- **Cắt ảnh theo khung dọc**
- **Thêm chữ "SQL Injection Attack" bằng Canva**

Tích hợp:

- **Slide mở đầu**
- **Slide cảnh báo**



3.4 Sử dụng Canva AI

Prompt 1:

Tạo layout video ngắn bằng tiếng Việt về chủ đề an ninh mạng, phong cách cảnh báo, nền tối, chữ rõ, có chỗ chèn ảnh và caption

Prompt 2:

Điều chỉnh layout để phù hợp video dọc (TikTok style), mỗi slide không quá 15 từ

Kết quả:

- **Layout rõ ràng, dễ đọc**

Chỉnh sửa:

- **Giảm chữ**
- **Tăng khoảng trắng**
- **Thêm hiệu ứng chuyển cảnh**

Tích hợp:

- **Xuất video hoàn chỉnh**



3.5 Demo SQL Injection

Prompt:

Cho ví dụ đơn giản về SQL Injection có thể bypass login

Output:

' OR '1'='1

Chỉnh sửa:

- **Tạo demo login form local**
- **Hiển thị rõ input và kết quả**

Tích hợp:

- **Đoạn cao trào video**

4. So sánh các công cụ AI

ChatGPT thể hiện khả năng vượt trội trong việc xây dựng nội dung có cấu trúc và logic rõ ràng. Trong khi đó, Google Gemini phù hợp hơn với vai trò phản biện và cải thiện nội dung.

Ví dụ cụ thể:

- **ChatGPT tạo kịch bản hoàn chỉnh ngay từ đầu**
- **Gemini phát hiện thiếu sót và đề xuất cải thiện**

DALL·E mạnh ở việc tạo hình ảnh nhanh, nhưng không đảm bảo độ chính xác kỹ thuật. Hình ảnh cần được chỉnh sửa lại để phù hợp ngữ cảnh.

Canva không tạo nội dung mạnh nhưng là công cụ tốt nhất để tích hợp tất cả thành sản phẩm hoàn chỉnh.

Kết luận:

- **ChatGPT: tạo nội dung**
- **Gemini: kiểm tra và cải thiện**

- **DALL·E: minh họa**
- **Canva: hoàn thiện**

5. Phân tích vai trò của AI

AI giúp giảm đáng kể thời gian sản xuất nội dung. Các công đoạn như viết kịch bản, tạo hình ảnh, thiết kế layout đều có thể hoàn thành nhanh chóng.

Tuy nhiên, AI có hạn chế nghiêm trọng:

- **Không kiểm chứng thông tin**
- **Có thể tạo nội dung sai nhưng hợp lý**
- **Không hiểu bối cảnh bảo mật**

Trong trường hợp SQL Injection, nếu lập trình viên sử dụng code AI mà không kiểm tra, hệ thống có thể bị tấn công dễ dàng.

AI thay đổi quy trình làm việc:

- **AI: tạo nhanh**
- **Con người: kiểm tra, quyết định**

Vấn đề đạo đức:

- **Không được phụ thuộc hoàn toàn vào AI**
- **Người dùng phải chịu trách nhiệm**
- **AI không chịu hậu quả**

6. Kết luận

Dự án chứng minh rằng AI là công cụ mạnh nhưng không thay thế được con người. Việc kết hợp nhiều công cụ AI giúp tăng hiệu suất, nhưng chất lượng cuối cùng phụ thuộc hoàn toàn vào khả năng kiểm soát và đánh giá của người sử dụng.

Sai lầm lớn nhất không phải là dùng AI, mà là tin AI mà không kiểm chứng.