

# AI-driven Virtual Storyteller

## KẾ HOẠCH TUẦN 1

Dự án: AI-driven Virtual Storyteller

Mã số: 242B1MDCP01

Thời gian: Tuần 1 (Khởi động dự án)

### I. MỤC TIÊU CHÍNH

Tuần đầu tiên đóng vai trò định hướng cho toàn bộ dự án, với các mục tiêu cụ thể như sau:

- Khảo sát các tài liệu học thuật, báo cáo nghiên cứu, và giải pháp mã nguồn mở liên quan đến các module cốt lõi của hệ thống kể chuyện tự động bằng AI.
- Phân tích khả năng tích hợp các công nghệ hiện có để tạo ra một pipeline kể chuyện hoàn chỉnh: từ việc sinh nội dung truyện, chuyển thành giọng đọc có cảm xúc, cho đến việc tạo hình ảnh minh họa phù hợp.
- Thực hiện một số thử nghiệm nhỏ để kiểm tra tính khả thi kỹ thuật của các hướng tiếp cận.
- Thiết lập nền tảng kỹ thuật, chuẩn bị cho việc triển khai thực tế từ tuần thứ hai trở đi.

### II. CÁC HẠNG MỤC NGHIÊN CỨU

#### 1. Tổng quan kiến trúc hệ thống kể chuyện bằng AI

Hệ thống kể chuyện tự động cần bao gồm các thành phần sau:

- NLP (Natural Language Processing):** Tạo ra truyện từ đề bài đầu vào hoặc gợi ý, đảm bảo yếu tố mạch lạc, cảm xúc và phân cảnh rõ ràng.
- TTS (Text-to-Speech):** Chuyển nội dung văn bản thành giọng đọc, có khả năng truyền tải cảm xúc (ví dụ: buồn, vui, hồi hộp).
- Visual Generator:** Sinh ảnh minh họa cho từng cảnh hoặc đoạn trong truyện, sử dụng các mô hình text-to-image (như Stable Diffusion, DALL·E 3).
- API xử lý và tích hợp:** Cầu nối giữa người dùng và các mô-đun AI, đồng thời quản lý dữ liệu đầu vào và đầu ra.

Trong tuần này, em tập trung vào việc  **nghiên cứu các thành phần trên về mặt lý thuyết và công nghệ hiện có**, thay vì phát triển tính năng cụ thể.

---

## 2. Nghiên cứu các tài liệu khoa học về mô hình sinh truyện (NLP)

### Nội dung khảo sát:

- **Các mô hình tạo nội dung tự động:** GPT-3.5/4, T5, LLaMA, StoryWriter, CTRL, và các phiên bản fine-tuned dành riêng cho storytelling.
- **Kỹ thuật điều khiển nội dung sinh:** prompt engineering, keyword conditioning, sentiment control, plot structuring.
- **Tổng hợp các nghiên cứu liên quan** từ các hội thảo như ACL, EMNLP, NAACL... về "story generation", "narrative generation", "multi-turn storytelling".

### Mục tiêu thử nghiệm nhỏ:

- Viết 2–3 prompt mẫu bằng tiếng Việt/Anh để kiểm tra khả năng sinh truyện của các mô hình hiện có (qua API OpenAI hoặc mô hình mã nguồn mở).
  - Đánh giá kết quả theo tiêu chí: logic cốt truyện, tính biểu cảm, khả năng chia đoạn hoặc phân cảnh.
- 

## 3. Khảo sát các công nghệ TTS có cảm xúc

### Nội dung khảo sát:

- So sánh các giải pháp chuyển văn bản thành giọng đọc hỗ trợ cảm xúc:
  - **ElevenLabs** (natural & emotional voices)
  - **Coqui TTS** (mã nguồn mở, có khả năng tùy chỉnh speaker)
  - **Google Cloud TTS** (multi-language support + emotional tuning)
- Tìm hiểu các kỹ thuật training voice model riêng (nếu cần về sau).

### Mục tiêu thử nghiệm nhỏ:

- Nhập một đoạn truyện ngắn (~3–4 câu) vào các hệ thống TTS trên để đánh giá chất lượng: độ rõ, độ tự nhiên, và cảm xúc thể hiện.
  - Phân tích xem có thể tích hợp API vào hệ thống sau này không.
- 

## 4. Khảo sát công nghệ sinh ảnh từ văn bản (Text-to-Image)

### Nội dung khảo sát:

- Tìm hiểu khả năng sinh ảnh minh họa từ mô tả văn bản sử dụng các mô hình như:
  - **Stable Diffusion v2/v3**: mã nguồn mở, hỗ trợ điều chỉnh nội dung ảnh.
  - **DALL·E 3 (OpenAI)**: sinh ảnh chất lượng cao theo prompt phức tạp.
  - **Midjourney, Leonardo.AI**: dùng để tham khảo về mặt thẩm mỹ và phong cách.
- Phân tích tính kiểm soát được trong quá trình sinh ảnh (chủ thể, cảnh vật, biểu cảm).

### Mục tiêu thử nghiệm nhỏ:

- Viết 1–2 đoạn miêu tả cảnh truyện ngắn, chạy prompt trên một mô hình để đánh giá chất lượng hình ảnh minh họa.
- Ghi nhận mức độ phù hợp giữa nội dung ảnh và văn bản đầu vào.

---

## 5. Tổng hợp tài liệu & chuẩn bị cho giai đoạn phát triển

- Ghi chú toàn bộ tài liệu đã đọc, link tham khảo, đánh giá ưu/nhược điểm từng mô hình hoặc dịch vụ.
- Lập bảng tổng hợp công nghệ phù hợp cho từng module (theo tiêu chí: chất lượng, tài nguyên yêu cầu, khả năng tùy chỉnh, chi phí).
- Nếu còn thời gian, có thể:
  - Khởi tạo repo dự án
  - Tạo sơ đồ kiến trúc ban đầu (bằng draw.io hoặc diagrams.net)
  - Lên danh sách các mô hình cần cài đặt thử trong tuần 2

---

## III. KẾT QUẢ DỰ KIẾN CUỐI TUẦN

- **Tài liệu tổng hợp**: Liệt kê các công nghệ và tài liệu học thuật quan trọng cho từng phần: NLP, TTS, Visual.
- **Bảng đánh giá**: Ưu/nhược điểm các giải pháp kỹ thuật hiện có, đề xuất hướng triển khai cụ thể cho tuần tiếp theo.
- **Một vài demo nhỏ**:
  - Một đoạn truyện được sinh tự động từ prompt
  - Một giọng đọc mẫu từ đoạn văn bản
  - Một hình ảnh minh họa sinh từ mô tả cảnh truyện
- **Kiến trúc sơ khai**: Phác thảo quy trình hoạt động của hệ thống kể chuyện tự động (pipeline tổng thể).

## IV. GHI CHÚ

- Giai đoạn này chưa tập trung viết code hoặc xây dựng hệ thống đầy đủ.
- Các thử nghiệm chỉ mang tính chất minh họa, mục tiêu là **định hướng kỹ thuật và chọn công cụ phù hợp** cho phát triển sau này.
- Kết quả thu được trong tuần này sẽ làm cơ sở cho việc lên kế hoạch chi tiết cho tuần 2 (bắt đầu phát triển từng module riêng lẻ).