**Earthship AI - Tài liệu hướng dẫn sử dụng (Tiếng Việt)**

**1. Cách sử dụng (cho người dùng thông thường):**

* **Bắt đầu:**
  1. **Khởi chạy ứng dụng:** Chạy tập lệnh Python. Thao tác này sẽ mở giao diện web trong trình duyệt mặc định của bạn (thường tại địa chỉ http://127.0.0.1:7871).
  2. **Đăng nhập hoặc tạo tài khoản:**
     + **Đăng nhập:** Nếu bạn đã có tài khoản, hãy nhập tên người dùng và mật khẩu vào phần "Đăng nhập" và nhấp vào "Đăng nhập".
     + **Tạo tài khoản:** Nếu bạn là người mới, hãy nhập tên người dùng và mật khẩu, sau đó nhấp vào nút "Tạo người dùng mới".
* **Trò chuyện với AI:**
  1. **Chọn tính cách:** Chọn một tính cách AI từ trình đơn thả xuống "Chọn tính cách AI". Mỗi tính cách có một phong cách và kiến thức khác nhau.
  2. **Chọn mô hình:** Chọn một mô hình AI từ trình đơn thả xuống "Chọn mô hình AI".
  3. **Bắt đầu cuộc trò chuyện mới:** Nhấp vào nút "Bắt đầu cuộc trò chuyện mới" để bắt đầu một cuộc trò chuyện mới.
  4. **Nhập tin nhắn của bạn:** Nhập tin nhắn của bạn vào hộp văn bản "Nhập tin nhắn của bạn".
  5. **Gửi tin nhắn:** Nhấn Enter hoặc nhấp vào nút "Gửi".
  6. **Xem phản hồi:** Phản hồi của AI sẽ xuất hiện trong cửa sổ trò chuyện. Phản hồi sẽ được truyền phát từng từ.
* **Các tính năng khác:**
  1. **Dừng tạo phản hồi:** Nhấp vào "Dừng tạo câu trả lời" để ngăn AI tạo thêm văn bản.
  2. **Sử dụng câu lệnh tạo sẵn:** Nhấp vào các nút trong phần "Thư viện công cụ" để chèn các câu lệnh đã viết sẵn vào tin nhắn của bạn.
  3. **Tải các cuộc trò chuyện trước đó:** Chọn một cuộc trò chuyện từ trình đơn thả xuống "Lịch sử trò chuyện" và nhấp vào "Load Chat" để xem các cuộc trò chuyện trước đây.
  4. **Đổi tên cuộc trò chuyện:** Tải một cuộc trò chuyện, nhấp vào "Rename Chat", nhập tên mới vào hộp văn bản xuất hiện và nhấn Enter.
  5. **Tìm kiếm trên Internet (Tùy chọn):** Chọn hộp "Sử dụng Internet để tìm kiếm" để cho phép AI tìm kiếm thông tin liên quan đến truy vấn của bạn trên web. Lưu ý: AI có thể sử dụng các liên kết cụ thể hoặc tìm kiếm web chung tùy thuộc vào tính cách đã chọn.
  6. **Trợ giúp về câu lệnh:** Nhấp vào nút "💡" bên cạnh đầu vào tin nhắn để nhận các đề xuất về cách cải thiện câu lệnh của bạn.
* **Hồ sơ người dùng:**
  1. **Chỉnh sửa hồ sơ:** Nhấp vào "Chỉnh sửa thông tin cá nhân", nhập mật khẩu của bạn khi được nhắc, sau đó điền vào các trường (Họ và tên, Tuổi, Giới tính, Chiều cao, Cân nặng, v.v.) và nhấp vào "Lưu thông tin".
  2. **Ẩn hồ sơ:** Nhấp vào "Ẩn thông tin cá nhân" để ẩn chi tiết hồ sơ của bạn khỏi giao diện trò chuyện.

**2. Cài đặt nâng cao (cho người dùng thành thạo):**

* **Truy cập cài đặt nâng cao:**
  + Hầu hết các cài đặt nâng cao không được hiển thị trực tiếp trong giao diện người dùng. Bạn sẽ cần sửa đổi mã hoặc tệp cấu hình để thay đổi chúng.
* **Các cài đặt chính:**
  + **Max Tokens (Độ dài phản hồi):**
    - **Vị trí:** Trong hàm chat\_with\_model.
    - **Cách thực hiện:** Thay đổi đối số max\_tokens (mặc định là 1000) khi gọi ollama.chat.
    - **Ngoài ra:** Trong hàm stream\_chat, bên trong options cho ollama.chat, bạn có thể điều chỉnh:
      * num\_predict: Kiểm soát độ dài phản hồi tối đa (tương tự như max tokens).
      * temperature: Điều chỉnh độ ngẫu nhiên (cao hơn = sáng tạo hơn).
      * top\_k: Giới hạn lựa chọn từ trong top K từ có khả năng xảy ra cao nhất.
      * top\_p: Giới hạn lựa chọn từ dựa trên xác suất tích lũy.
  + **Max Input (Độ dài đầu vào tối đa):**
    - **Vị trí:** Không có độ dài đầu vào tối đa nghiêm ngặt nào được thực thi bởi chính mã. Tuy nhiên, đầu vào quá dài có thể gây ra sự cố với mô hình AI hoặc giao diện.
    - **Cách thực hiện:** Hộp văn bản msg trong giao diện Gradio có max\_length là 1000. Bạn có thể sửa đổi thông số này nếu cần. Hãy lưu ý đến các giới hạn của mô hình AI cơ bản.
  + **Tùy chọn hiển thị:**
    - **Vị trí:** Giao diện Gradio (gr.Chatbot, gr.Textbox, v.v.) và mã CSS xác định hầu hết các khía cạnh trực quan.
    - **Cách thực hiện:**
      * Sửa đổi các tham số height, render, elem\_classes, v.v. của các thành phần Gradio.
      * Chỉnh sửa CSS trong phần with gr.Blocks(css=...) để thay đổi kiểu dáng của các phần tử.
  + **Tốc độ tạo văn bản (Độ trễ):**
    - **Vị trí:** Tốc độ truyền phát được xác định bởi tốc độ tạo văn bản của mô hình Ollama và cách hàm stream\_chat tạo ra các khối.
    - **Cách thực hiện:**
      * Không có độ trễ tích hợp. Bạn có thể thêm độ trễ nhân tạo bằng cách sử dụng time.sleep() trong vòng lặp stream\_chat, nhưng điều này có thể làm cho giao diện có cảm giác không phản hồi.
      * Yếu tố chính là tốc độ xử lý của mô hình Ollama.
  + **Thay đổi câu lệnh và tính cách:**
    - **Vị trí:** Tệp prompts.txt.
    - **Cách thực hiện:**
      * Chỉnh sửa prompts.txt để sửa đổi EXAMPLE\_RESPONSES, PERSONALITIES và PREMADE\_PROMPTS.
      * Tuân theo định dạng hiện có (các từ điển với các khóa cụ thể như "system", "user", "links").
      * Khởi động lại ứng dụng để các thay đổi có hiệu lực.
  + **Thay đổi mô hình:**
    - **Vị trí:** Các từ điển MODEL\_DISPLAY\_NAMES và AVAILABLE\_MODELS.
    - **Cách thực hiện:**
      * Thêm hoặc sửa đổi các mục trong các từ điển này, ánh xạ tên hiển thị với số nhận dạng mô hình Ollama.
      * Đảm bảo các mô hình bạn thêm đã được cài đặt và có thể truy cập được đối với Ollama.
      * Khởi động lại ứng dụng để các thay đổi có hiệu lực.
  + **Dữ liệu người dùng:**
    - **Vị trí:** Thư mục userdata.
    - **Cách thực hiện:**
      * Dữ liệu người dùng được lưu trữ trong các tệp JSON bên trong các thư mục con được đặt tên theo từng tên người dùng.
      * Bạn có thể trực tiếp kiểm tra hoặc sửa đổi các tệp này, nhưng hãy cẩn thận để không làm hỏng cấu trúc.
  + **Bảng điều khiển quản trị (Admin Panel):**
    - **Vị trí:** Mã liên quan đến admin\_panel, user\_selector, admin\_chatbot, v.v.
    - **Cách thực hiện:**
      * Bảng điều khiển quản trị có thể truy cập được đối với người dùng "admin" theo mặc định.
      * Bạn có thể tùy chỉnh các tính năng quản trị bằng cách sửa đổi các hàm và thành phần Gradio có liên quan.
  + **Ghi nhật ký (Logging):**
    - **Vị trí:** Dòng logging.basicConfig.
    - **Cách thực hiện:** Thay đổi level để kiểm soát mức độ chi tiết của nhật ký (ví dụ: logging.DEBUG để có nhật ký chi tiết hơn). Bạn cũng có thể thay đổi format của các thông báo nhật ký.

**Lưu ý quan trọng (cho người dùng nâng cao):**

* **Ollama:** Đảm bảo bạn đã cài đặt và chạy Ollama đúng cách. Bạn sẽ cần tải xuống các mô hình AI mà bạn muốn sử dụng trong Ollama.
* **Khởi động lại:** Sau khi thực hiện thay đổi mã, bạn thường cần khởi động lại tập lệnh Python để chúng có hiệu lực.
* **Xử lý lỗi:** Mã có một số xử lý lỗi cơ bản, nhưng bạn có thể muốn thêm xử lý lỗi mạnh mẽ hơn, đặc biệt nếu bạn đang thực hiện các sửa đổi đáng kể.
* **Bảo mật:** Nếu bạn dự định triển khai ứng dụng này một cách công khai, hãy xem xét cẩn thận các vấn đề bảo mật, đặc biệt là liên quan đến dữ liệu người dùng và xác thực.