**📄 DỰ ÁN: News2Threads**

**🌟 MỤC TIÊU DỰ ÁN**

Xây dựng một hệ thống tự động thu thập tin tức từ trang [Báo Mới (baomoi.com)](https://baomoi.com/), tóm tắt nội dung bằng AI (GPT) và đăng lại nội dung ngắn gọn lên Threads một cách tự động.

**📅 NỘI DUNG DỰ ÁN**

1. **Thu thập tin tức theo thời gian thực** từ trang Baomoi.com theo từ khoá/tag
2. **Lưu trữ bài viết tạm thời** trong cơ sở dữ liệu
3. **Trích xuất nội dung chính và hình ảnh** của bài viết
4. **Sử dụng GPT** để tóm tắt nội dung trong 2-3 câu ngắn
5. **Tự động đăng nội dung tóm tắt lên Threads** (dùng Playwright)

**✅ YÊU CẦU & CÔNG NGHỌ DÙNG**

**1. Ngôn ngữ chính: Python 3.10+**

**2. Công nghệ / Thư viện:**

| **Mục đích** | **Công cụ** |
| --- | --- |
| Crawling | requests, BeautifulSoup, hoặc Playwright |
| Tóm tắt AI | openai SDK (GPT-3.5/GPT-4) |
| DB tạm | SQLite3 hoặc PostgreSQL |
| Automation | Playwright (Python) |
| Scheduler | schedule hoặc crontab |

**3. Cài đặt phụ thuộc:**

pip install openai playwright beautifulsoup4 schedule

playwright install

**🔄 FLOW HOẠT ĐỘNG HỆ THỐNG**

1. Truy cập trang tag (VD: https://baomoi.com/da-nang-tag11450.epi)

↓

2. Lấy danh sách bài mới → Truy cập link chi tiết bài viết

↓

3. Trích xuất: tiêu đề, đoạn văn chính, hình ảnh

↓

4. Gửi nội dung qua GPT → nhận tóm tắt 2-3 câu

↓

5. Lưu bài viết và tóm tắt vào DB

↓

6. Mở Threads qua Playwright → Tự động đăng nội dung + ảnh

**🔗 Mở rộng (Tùy chọn):**

* Thêm giao diện dashboard (Streamlit / Flask)
* Phân loại bài viết theo chủ đề (AI)
* Lọc bài spam, trùng nội dung
* Lịch đăng định kì vào khung giờ hot

**📊 KPI/Đầu ra:**

* 10~20 bài/ngày được tóm tắt và đăng lại
* Lưu trữ dữ liệu vào DB để theo dõi
* Log thời gian crawl, tóm tắt, đăng bài

💼 **Người thực hiện:** [Tên bạn]

📅 **Thời gian khởi tạo:** [Ngày bắt đầu]

🎉 **Trạng thái:** ✨ Ready to build!