**📑 BẢN GHI NHỚ & BÁO CÁO TIẾN ĐỘ**

**Dự án:** AI Translator Agent – Dịch Tài Liệu Lớn  
**Ngày lập:** 25 / 05 / 2025  
**Người thực hiện:** ChatGPT (hỗ trợ)

**1. THÀNH QUẢ ĐÃ HOÀN THÀNH**

| **Mảng** | **File / Thành phần** | **Trạng thái** | **Ghi chú** |
| --- | --- | --- | --- |
| **Cấu hình mô hình** | model\_config.py | ✔️ | 3 cấp độ (thô/chuẩn/cao cấp) – map alias – tên model **hợp lệ** (gpt‑3.5‑turbo / gpt‑4o / gpt‑4.1) |
| **Định tuyến (đồng bộ)** | model\_router.py | ✔️ | Fallback đa‑key ‑ đa‑provider, xử lý lỗi KEY/PARSE/FALLBACK, hỗ trợ MOCK\_MODE |
| **Định tuyến (async)** | model\_router\_async.py | ✔️ | Logic tương tự bản sync, benchmark & logging đầy đủ |
| **Logging** | logger\_utils.py | ✔️ | failed\_model.log, benchmark.log + tag lỗi chi tiết |
| **Giao diện** | app.py | ✔️ | 2 tab (*Dịch tài liệu* & *Dashboard*), xuất Word + PDF, xử lý overlap, cache kết quả |
| **Dashboard** | trong app.py | ✔️ | Biểu đồ hiệu suất, bảng lỗi gần nhất |
| **Mock mode** | biến môi trường | ✔️ | Test UI/CI mà không tốn token |
| **Injection API‑key** | Chèn key người dùng vào API\_KEYS | ✔️ | Tránh lỗi mọi key fail |

**2. VẤN ĐỀ ĐÃ GIẢI QUYẾT HÔM NAY**

1. **Lỗi NoneType** → đã bổ sung kiểm tra response.json() & log PARSE.
2. **Kiểu key rỗng** → lọc None trong api\_keys.py.
3. **Key từ UI không dùng** → chèn vào API\_KEYS ngay khi người dùng nhập.
4. **Không khớp cấp độ** → thêm STYLE\_LEVEL\_MAP.
5. **Tên model sai** → đồng bộ với danh sách /v1/models.

**3. DANH MỤC VIỆC CẦN LÀM NGÀY MAI (ƯU TIÊN ↓)**

| **Ưu tiên** | **Việc cần triển khai** | **Kết quả mong đợi** |
| --- | --- | --- |
| 🔴 | **01 – Tối ưu Prompt** (system/user) | Bản dịch mượt, ít sai ngữ cảnh |
| 🟠 | **02 – Chế độ A/B** (so sánh 2 mô hình) | Chọn mô hình tốt nhất cho từng tài liệu |
| 🟠 | **03 – Xuất & gửi log (Excel/email)** | Theo dõi chi phí + lỗi dài hạn |
| 🟡 | **04 – Docker hoá** | Dễ deploy server/CI |
| 🟢 | **05 – UI nâng cao** (Dark mode, toggle Mock) | Trải nghiệm người dùng |
| 🟢 | **06 – Phân quyền & lưu token người dùng** | Sẵn sàng đa user (SaaS) |

**Lưu ý:** Bắt đầu ngày mai ưu tiên “01 – Tối ưu Prompt”. Khi hoàn thành, chạy benchmark so sánh trước/sau để đo chất lượng.

**4. GHI CHÚ KHÁC**

* **Chi phí ước tính hiện tại:** ~0,5 USD / 1 k từ với GPT‑4o (mức *Chuẩn*).
* **Tệp log** đã kiểm chứng hiển thị tốt trên Dashboard.
* **Mock mode** vẫn bật bằng export MOCK\_MODE=True khi cần demo offline.