📌 Project Overview

Title: Modular AI Report Generation Backend with FastAPI + RAG + LangChain

This project delivers a production-ready backend system for generating dynamic reports using LLMs, Retrieval-Augmented Generation (RAG), and secure API orchestration. Built with a modular structure, the system is designed for scalability, traceability, and multilingual support.

🔧 Technology Stack

- **Backend Framework**: FastAPI (async-ready, modern, production-grade)
- **LLM Orchestration**: LangChain, RAG pattern integration
- **Vector DB**: FAISS + Custom Embedding pipelines
- **Security**: JWT Auth, API key protection
- **Logging**: English + Chinese multilingual log system
- **Testing**: pytest for API and inference coverage
- **Packaging**: Ready-to-deploy ZIP structure with startup scripts
- **Internationalization**: Dual-language log messages (en/zh), structured logger with contextual tracing

---

☑️ Key Features

- **Plug-and-Play Architecture**: Easily extensible for future agents, prompts, or workflows
- **Rapid Deployment**: System can be deployed and tested immediately upon extraction
- **Audit-Ready Logging**: Logs structured for production use, includes language, timestamp, log level, and source module
- **Clean API Design**: /inference, /health, /docs, and future agent endpoints

---

📦 Deliverables

1. Fully functional backend system (source code in structured modules)
2. Deployment guide (README.md included)
3. Final packaging (.zip) with all scripts and config

4. Sample test scripts and logging outputs
5. API key / JWT-based auth system, ready for integration

👩‍💻 Maintainer

The system is developed and maintained by a full-stack LLM engineer with experience in:

- Secure, multilingual API backends
- Retrieval-augmented generation for reports and summaries
- Custom logging frameworks and lifecycle management
- Upwork-based remote deployments and consulting

🌍 Deployment Context (Optional)

This system can be easily hosted on:

- VPS (Ubuntu/Windows)
- Cloud VM (AWS, GCP, Azure)
- Local test environments via .bat or shell scripts
- Container solutions (optional, though Docker is avoided as per client preference)