# 階段 2：Demo 系統落地每日任務細化版

🎯 目標：9/1 面試可現場操作完整 RAG + 多模態 + Live2D 多語助理

## 8/15（四） - 環境準備與資料集蒐集

* 安裝並配置 Python 環境（conda / poetry）
* 安裝必要套件（LangChain、FAISS / Milvus、LLaMA、TTS SDK、Live2D SDK）
* 蒐集並挑選 2-3 份 PDF 報告（產業 / 技術）做為 Demo 資料集
* 產出：資料夾結構 + 安裝紀錄

## 8/16（五） - 資料抽取與清理

* 用 PyMuPDF / pdfplumber 抽取 PDF 文字
* 若有圖片 → 用 OCR（PaddleOCR / Tesseract）提取文字
* 清理換行、頁碼、特殊字元
* 產出：乾淨的 .txt 或 .jsonl 檔

## 8/17（六） - Chunking 與 Embedding

* 使用 LangChain RecursiveCharacterTextSplitter 切 chunk（500 tokens + 50 overlap）
* 用 BGE / Instructor Embedding 生成向量
* 存入向量資料庫（FAISS 或 Milvus）
* 產出：可檢索的向量庫 v1

## 8/18（日） - Fine-tuning 準備（LoRA / QLoRA）

* 選擇基座模型（LLaMA 3.2 8B / 13B）
* 整理高品質 Q&A 訓練集（可先用 GPT-4 生成）
* 測試 LoRA 訓練流程（小批量）
* 產出：LoRA 訓練配置檔

## 8/19（一） - LoRA 微調與測試

* 正式用生醫 / 產業報告數據集進行微調
* 測試模型在封閉領域問答的準確率
* 產出：微調後的模型檔

## 8/20（二） - 建立 RAG API

* 用 FastAPI 寫檢索 + LLM 回答 API
* 輸入問題 → 檢索 top-k chunk → 模型生成 → 回傳答案
* 產出：可本地或雲端部署的 API

## 8/21（三） - 檢索優化

* 增加 ± chunk 上下文檢索
* Hybrid 檢索（向量 + 關鍵詞）
* 加入答案來源引用
* 產出：檢索準確度報告

## 8/22（四） - 多模態檢索 – 圖片與表格處理

* 用 LLaMA 3.2 Vision / BLIP-2 從 PDF 圖片抽取資訊
* 將圖片內容轉為向量並存入資料庫
* 產出：圖文混合檢索結果

## 8/23（五） - 多語 TTS 整合

* 測試 Azure / ElevenLabs / OpenAI TTS API
* 支援中文、英文、日文語音輸出
* 產出：多語音合成測試檔

## 8/24（六） - Live2D 模型準備

* 建立或導入 Live2D 模型（可用免費範例）
* 配置 Live2D SDK / VTube Studio API
* 產出：本地可渲染的 Live2D 角色

## 8/25（日） - Live2D + TTS 整合

* 接收 API 回覆 → TTS 語音 → Live2D 嘴型同步
* 產出：Live2D 說話 Demo v1

## 8/26（一） - 多語 Live2D

* 增加語言切換功能（UI / 按鍵觸發）
* 產出：多語 Live2D 說話 Demo

## 8/27（二） - RAG + Live2D 全流程串接

* 問題輸入 → RAG 檢索 → LLM 回答 → 多語 Live2D 說話
* 產出：第一版完整系統 Demo

## 8/28（三） - 多模態 Live2D 整合

* 支援用圖片/表格做檢索並口頭回答
* 產出：多模態 Live2D Demo

## 8/29（四） - 壓測與錄影

* 模擬面試問答，錄製 Demo 影片做備用
* 產出：Demo 錄影檔

## 8/30（五） - Bug 修正 & UI 美化

* 確保面試現場操作流暢
* 產出：最終版 Demo

## 8/31（六） - 面試預演

* 全程走一遍流程，控制在 5~7 分鐘內完成
* 產出：面試演練紀錄