AI-Powered Resume Scanner - Step-by-Step Learning Plan

This roadmap is designed to teach you AI development by **building** an AI-powered resume scanner step by step. Instead of blindly copy-pasting code, you'll work through focused exercises and mini-projects to master PDF parsing, text cleaning, GPT integration, and Gradio UI development.

Phase 1 — PDF Parsing (Day 1)

1. Understand how PDFs store text and why parsing is tricky.
2. Learn PyMuPDF (`fitz`) and pdfminer, compare their outputs.
3. Write a function to extract text from one PDF.
4. Exercise: Parse 3 different resumes and print extracted text.

Phase 2 — Text Cleaning & Preprocessing (Day 2)

1. Learn regular expressions to clean up noisy data.
2. Remove extra spaces, symbols, and stopwords using NLTK.
3. Exercise: Compare before/after cleaning on 3 resumes.

Phase 3 — GPT Integration (Day 3)

1. Set up OpenAI API and test a basic prompt.
2. Send cleaned resume text + job description to GPT for scoring.
3. Understand prompt engineering and experiment with responses.
4. Exercise: Change prompts to generate resume summaries instead.

Phase 4 — Gradio UI (Day 4)

1. Learn Gradio basics: inputs, outputs, buttons, event handling.
2. Make a simple app with a textbox and output field.
3. Integrate PDF upload, parsing, and GPT scoring step by step.
4. Exercise: Build a Gradio app that just previews extracted text first.

Phase 5 — Final Project Integration (Day 5+)

1. Combine PDF parsing, cleaning, GPT scoring, and Gradio UI.
2. Add error handling for missing files and invalid PDFs.
3. Experiment with bulk resume uploads and CSV exports.
4. Checkpoint: Rebuild a minimal version of the app from scratch without tutorials.

Pro Tip: Don't rush. Work on one concept at a time. The goal is **deep understanding**, not just getting the app to run.