Project Roadmap: "Shell Company Sherlock"

AI-Powered Fraud Investigation System using LangGraph

Phase 1: Architecture & State Design (The Foundation)

Goal: Define how the agents "talk" and what data moves through the system.

1. **Define the "Case File" (Shared State):**
    - Design the data structure that will pass between agents. This acts as the memory for the investigation.
    - **Required Fields:**
        - Company_Name (Input)
        - Risk_Score (Integer, 0-100)
        - Evidence_Log (List of findings/notes)
        - Investigation_Status (e.g., "Pending", "Escalated", "Cleared")
        - Human_Analyst_Feedback (For the Human-in-the-Loop stage)
2. **Map the Agent Personas:**
    - **Agent A (The Registrar Scout):** Dedicated to verifying official existence (incorporation dates, business licenses).
    - **Agent B (The Geo-Spatial Analyst):** Dedicated to verifying physical presence (maps, street view, zoning).
    - **Agent C (The Digital Footprint Tracer):** Dedicated to reputation (social media, news mentions, CEO verification).
    - **The Supervisor (Router):** The decision-maker that reviews the total Risk Score and routes the next step.

Phase 2: The Tool Belt (The Capabilities)

Goal: Give your agents the ability to interact with the outside world. *Note: Start by creating "Mock Tools" (fake functions that return hardcoded data) so you can test the logic without spending money on API credits.*

1. **Develop the "Registry Lookup" Tool:**

- **Input:** Company Name.
- **Output:** Incorporation Date, Registered Address, Director Names.
- *Real Implementation:* Connect to OpenCorporates API or a government business registry scraper.

2. **Develop the "Location Verifier" Tool:**
    - **Input:** Address string.
    - **Output:** Location Type (Commercial/Residential), GPS Coordinates.
    - *Real Implementation:* Connect to Google Maps Geocoding API or Places API.
3. **Develop the "Web Search" Tool:**
    - **Input:** Search Query (e.g., "TechSolutions LLC scam reviews").
    - **Output:** Summarized search snippets.
    - *Real Implementation:* Connect to Tavily API or SerperDev (optimized for LLMs).

Phase 3: Agent Logic Construction (The Brains)

Goal: Program the decision-making rules for each specific agent.

1. **Build the Registrar Agent:**
    - Logic: If the "Incorporation Date" is less than 6 months ago → Add +20 to Risk Score.
    - Logic: If the Director's name appears in a "Banned Persons" list → Add +50 to Risk Score.
2. **Build the Geo-Spatial Agent:**
    - Logic: If the address type is "Residential" or "P.O. Box" → Add +30 to Risk Score.
    - Logic: If multiple different companies share the exact same address → Flag as "Suspected Shell Farm."
3. **Build the Supervisor Agent:**
    - Logic: Review the cumulative Risk_Score.

- **Routing Rule:**
    - Score > 75: Route to "Freeze Account."
    - Score 30-75: Route to "Human Review Queue."
    - Score < 30: Route to "Clear Transaction."

Phase 4: LangGraph Orchestration (The Assembly)

Goal: Connect the agents into a cyclic graph workflow.

1. **Initialize the StateGraph:**
    - Set up the graph using your defined "Case File" schema.
2. **Add Nodes:**
    - Register your Agent Logic functions (from Phase 3) as nodes in the graph.
3. **Define Edges (The Flow):**
    - Connect the "Start" point to the Registrar Agent.
    - Connect the Registrar Agent to the Geo-Spatial Agent (or run them in parallel).
    - Connect all analysis agents to the Supervisor.
4. **Implement Conditional Logic:**
    - Program the Supervisor node to dynamically choose the next path based on the Risk Score. This is the "Conditional Edge."

Phase 5: Safety & Human-in-the-Loop (The Enterprise Grade)

Goal: Ensure the AI doesn't make dangerous autonomous decisions.

1. **Implement interrupt_before:**
    - Configure LangGraph to strictly pause execution before the final "Freeze Account" action.
    - This forces the system to wait for a human signal (e.g., typing "Approved" in the console) before proceeding.

2. **Memory Persistence (Checkpointer):**
    - Add a checkpointer (like SQLite or simple memory saver) to the graph.
    - *Why?* This allows the investigation to "sleep" while waiting for a human manager to review it, and "wake up" exactly where it left off days later.

Phase 6: Interface & Deployment (The Showcase)

Goal: Create a visual way to demonstrate the project.

1. **Build a Simple UI (Streamlit):**
    - Create a text box for "Enter Company Name."
    - Create a "Run Investigation" button.
    - Display the "Case File" updating in real-time as agents finish their tasks.
    - **Crucial:** Add a "Manager Approval" button that triggers the resume function for the Human-in-the-Loop step.
2. **Final Validation:**
    - Test with a known legitimate company (e.g., "Microsoft").
    - Test with a fake shell company (input data that triggers your red flags).