Gmail
AI AGENT
LangGraph

Overview

1. AI-Driven Prospecting

- Identifies global companies by service vertical using LLM suggestions
- Validates prospects through multi-source analysis (search APIs, web scraping)

2. Smart Contact Extraction

- Extracts professional emails with pattern matching and domain validation
- Filters out generic addresses (noreply, support) for high-quality leads

3. Automated Workflow Orchestration

- Uses LangGraph to manage stateful email sequences
- Routes responses intelligently (replies vs. automated messages)

4. Context-Aware Engagement

- Generates personalized emails with service-specific value propositions
- Maintains conversation history for coherent follow-ups

5. Self-Optimizing Infrastructure

- Auto-retries failed requests with rate limiting
- Adapts **to timestamps** and communication patterns

6. End-to-End Analytics

- Tracks lead sources, response rates, and conversion metrics
- Exports structured data (CSV) ready for CRM integration

Methodology

- **Data Pipeline**
- Sources: Web scraping, APIs, LLM suggestions
- Processing: Email/phone regex, domain validation, contact page detection
- **Key Variables**
- Raw: Company names, URLs, service types
- Derived: Contact accessibility, email quality score, service relevance
- **Technical Core**
- Hybrid rule-based/LLM validation
- LangGraph-managed workflow
- Concurrent processing with rate limiting
- Focused on reliable, actionable lead generation through multi-stage validation.

Framework / tools used

- **Core Frameworks & Libraries**

1. **LangGraph**

- Why: Manages stateful workflows (lead validation → email sequencing) with flexible DAG orchestration**.**

2. Llama-3 (via Groq API)

- Why: High-speed LLM for company analysis/email generation (70B params, low-latency).

3. BeautifulSoup + Requests

- Why: Lightweight HTML scraping/parsing for contact page detection.

4. Tavily API

- Why: Real-time web search for email/domain validation.
- Concurrent Processing

5. Tools: ThreadPoolExecutor, asyncio

- Why: Parallelize API calls/scraping for speed.

6. Regex + Fuzzy Matching

- Libraries: re, fuzzywuzzy
- Why: Robust email/phone extraction and deduplication.

Model Selection

- **Why use Ollama local model?**

1. Data Never Leaves Your Infrastructure

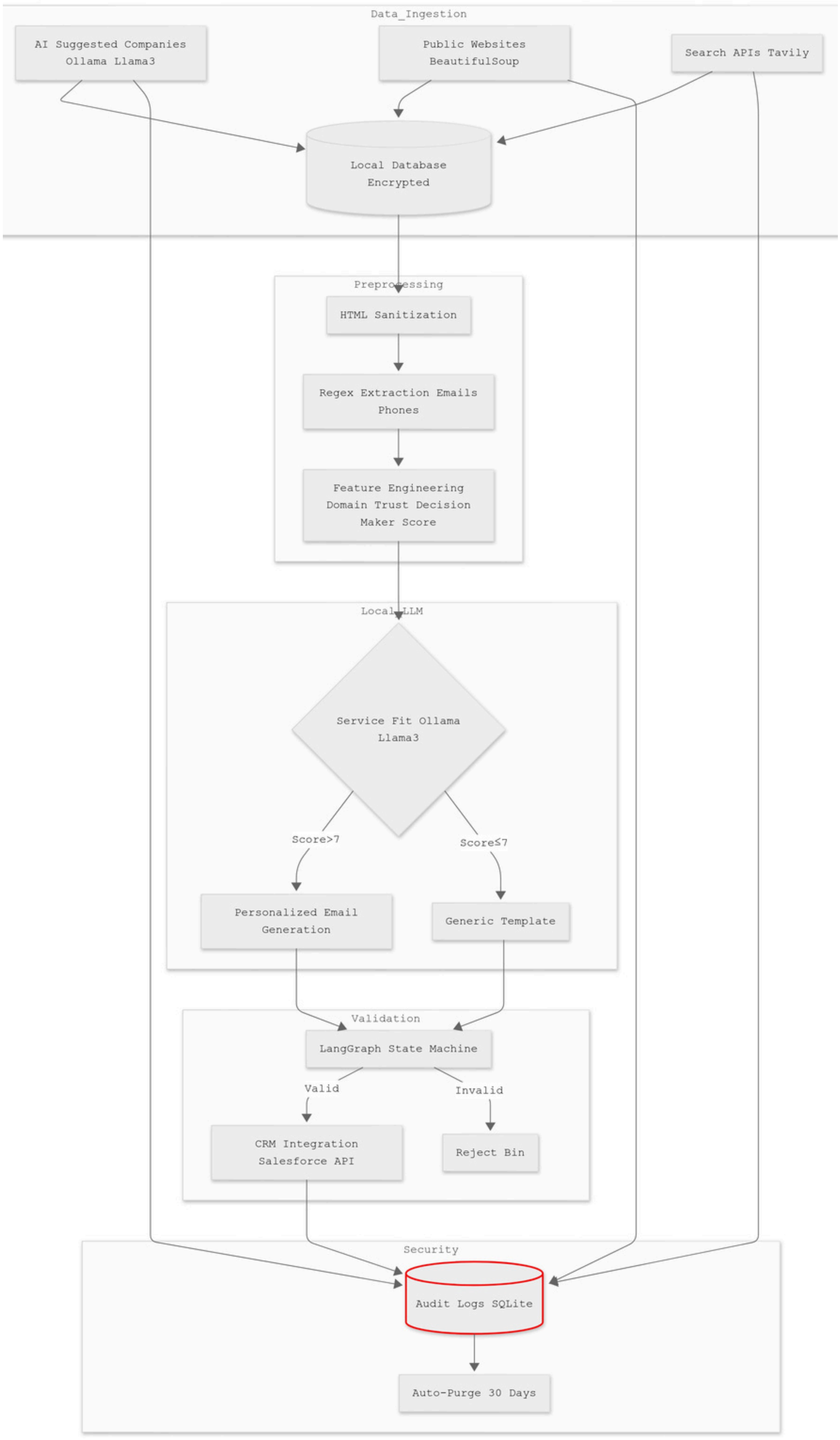
- Critical for: Sensitive lead data (company contacts, proprietary research)
- Avoids: Third-party API risks (e.g., OpenAI/Groq data logging)

2. Reduced Attack Surface

- **No external API = No exposure to:**
- Man-in-the-middle attacks
- Vendor data breaches

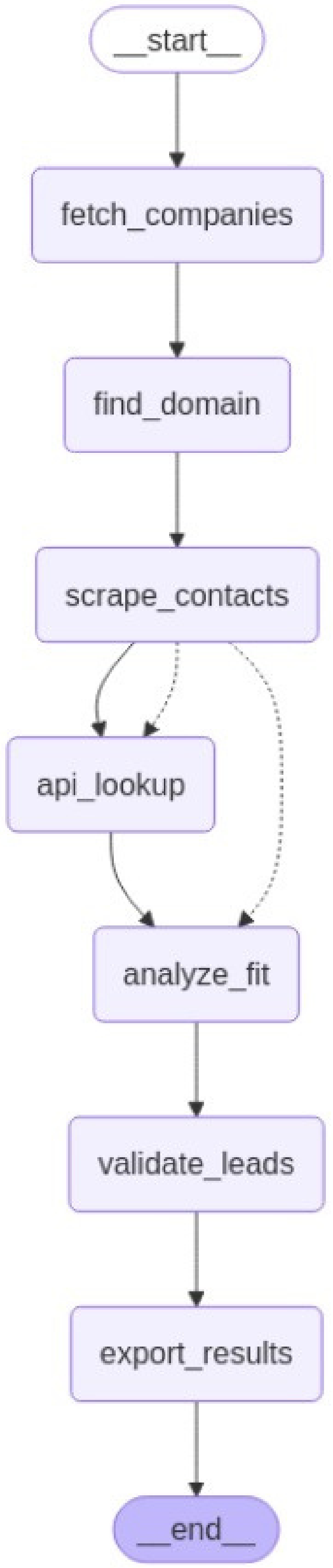
3. Use of quantized models (Llama-3-8B) for faster local inference

Model Architecture

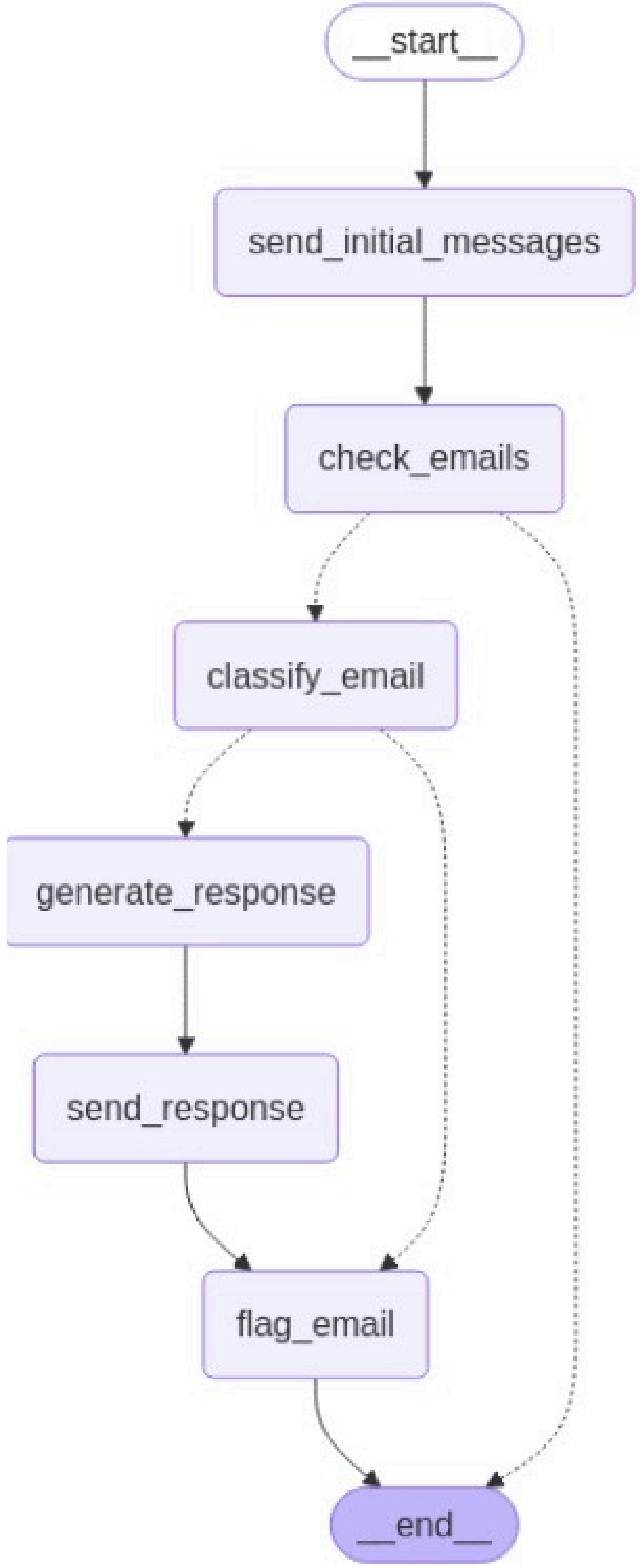

Data_Ingestion
AI Suggested Companies Ollama Llama3
Public Websites BeautifulSoup
Search APIs Tavily
Local Database Encrypted
Preprocessing
HTML Sanitization
Regex Extraction Emails Phones
Feature Engineering Domain Trust Decision Maker Score
Local_LLM
Service Fit Ollama Llama3
Score>7
Score≤7
Personalized Email Generation
Generic Template
Validation
LangGraph State Machine
Valid
Invalid
CRM Integration Salesforce API
Reject Bin
Security
Audit Logs SQLite
Auto-Purge 30 Days

workflow diagram

Lead Generation



Outreach



Future Improvements

- **Develop multi-channel social media agents to automate outreach workflows on LinkedIn and Twitter/X, expanding engagement beyond email.**
- **Enhance scraping algorithms with improved email extraction logic to boost the lead-to-email ratio and reduce missing contact data.**

Thank you!!