Lily Agent Documentation

Introduction

This project leverages FireCrawl and DeepSeek R1 to build an end-to-end AI agent for automated lead generation and outreach. With a user-friendly web interface, the system guides users through a multi-steps workflow (or a dashboard view) that gathers, processes, and prioritizes potential customer data before generating personalized outreach emails. The demo and documentation use DuPont Tedlar as an example.

Core Features

Fully Automated Workflow

The end-to-end process—from crawling data and processing content to prioritizing leads and drafting outreach emails—is fully automated.

High Extensibility & Reusability

The modular design and well-defined APIs allow the system to be easily adapted for any parent company's lead generation and outreach needs.

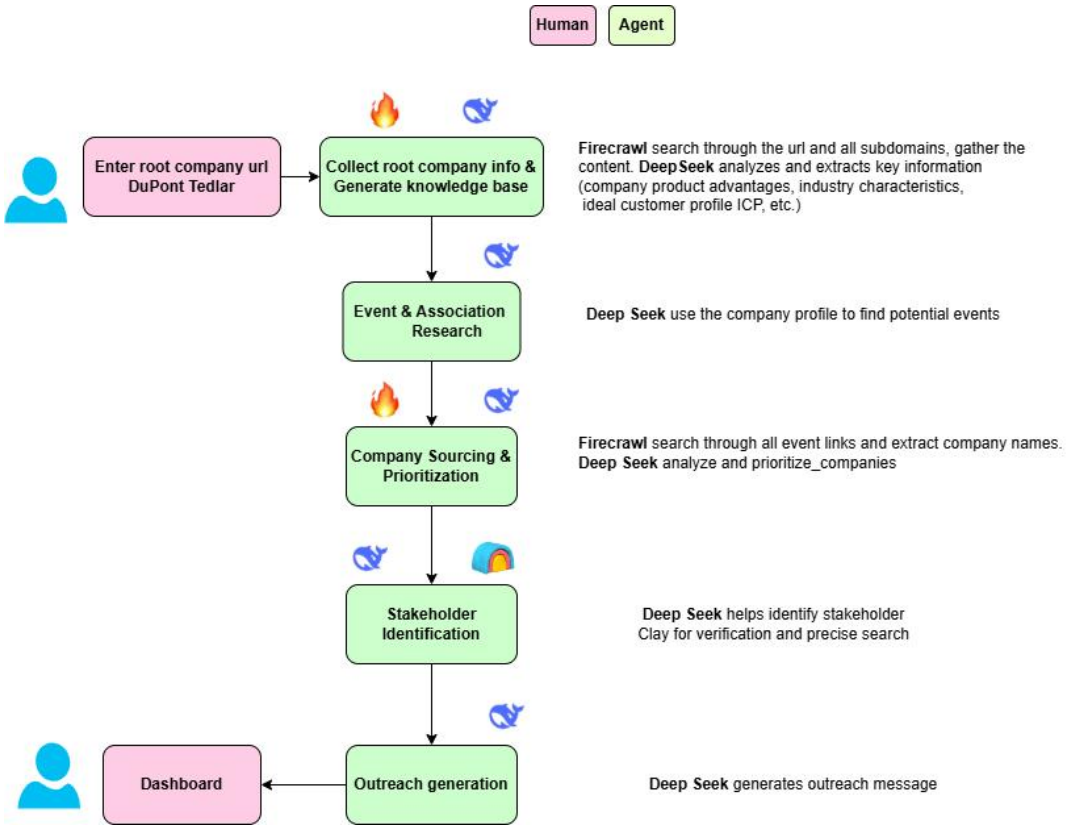
Web Interface

Offers both a step-by-step wizard and a dashboard view. Users can inspect intermediate outputs (e.g., potential events, raw customer lists, prioritized companies) through "View" buttons. In the dashboard, data is displayed in a clean, card-based layout, and each prioritized company card includes an "Outreach" button to generate and preview personalized emails.

Cost-Effective

The whole process from scratch to end costs $0.07 LLM API fees.

AI agent workflow



1. Enter Root Company URL (e.g., DuPont Tedlar)

The user provides the main or "root" company URL. This serves as the starting point for gathering essential data about the company—such as its history, product offerings, and value propositions.

2. Collect Root Company Info & Generate Knowledge Base

FireCrawl scrapes the provided URL (and subdomains, if needed) to collect website content. DeepSeek analyzes and extracts key information (e.g., product advantages, industry focus, ICP, etc.) to produce a structured knowledge base.

3. Event & Association Research

DeepSeek searches relevant public sources (industry portals, event listings) to compile potential trade shows, conferences, or associations where target customers might be found. DeepSeek also refines the raw data into a structured list of events and associations with links.

FireCrawl extracts company names from all event links in batch and generates structured list.

4. Company Sourcing & Prioritization

DeepSeek gathers information about potential customer companies (via public sites, directories, or references in the knowledge base). Ranking these companies based on criteria such as revenue, size, industry fit, and alignment with the root company's products or services. The output is a prioritized list (e.g., prioritized_companies.json).

5. Stakeholder Identification

DeepSeek identifies key decision-makers (stakeholders) within each prioritized company—locating names, titles, email addresses, and phone numbers.

For verification and enrichment, Clay APIs may be integrated and help for precise search.

6. Outreach Generation

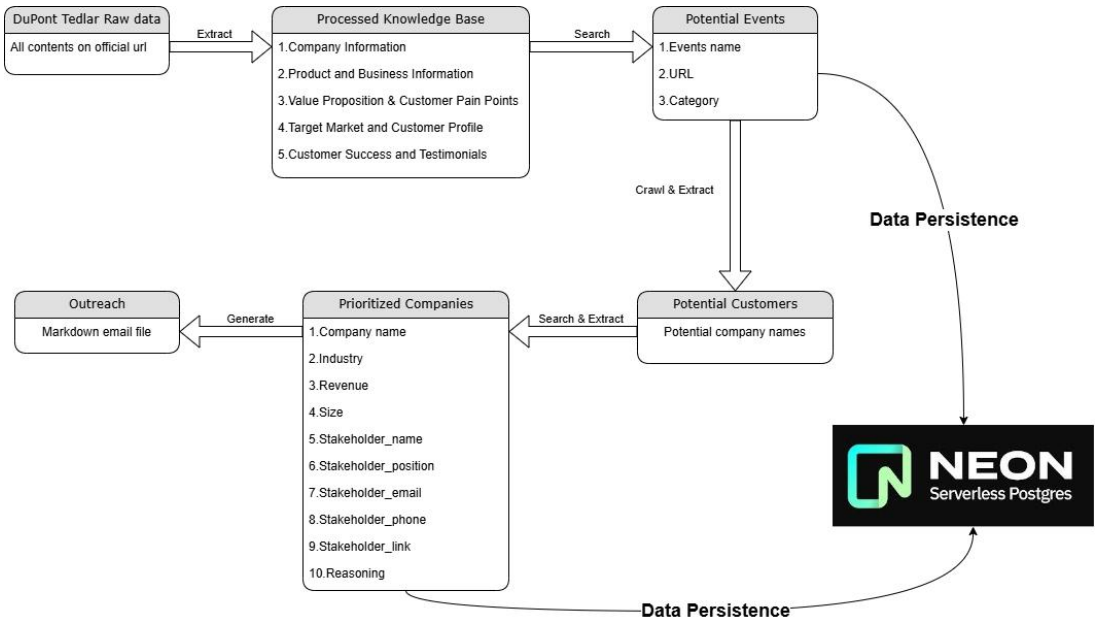
DeepSeek leverages the root company's knowledge base and stakeholder information to draft highly personalized outreach emails.

7. Dashboard

A unified interface consolidates the findings (events, potential customers, prioritized companies, and stakeholders). Users can view each category in a separate column, examine details, and perform outreach actions in one place.

These emails include a professional introduction, targeted product value statements, and a clear call to action—ready for the user to preview.

Data processing steps



The workflow begins by crawling the DuPont Tedlar website to extract raw content, which is then transformed into a structured knowledge base using DeepSeek. The system proceeds to derive three key outputs:

Potential Events

Extracted from public sources and processed by DeepSeek, these events (e.g., associations and exhibitions) are persisted in the database for long-term reference.

Potential Customers

A list of potential customer companies is compiled and temporarily cached. This list serves as a source for later processing.

Prioritized Companies

DeepSeek ranks the potential customers based on various criteria. The resulting prioritized list is also persisted in the database.

Finally, based on the prioritized companies, the system generates personalized outreach emails by combining the company profile with the stakeholder information.

This design ensures that critical data (potential events and prioritized companies) is stored permanently while the remaining artifacts are maintained as cache for flexibility and performance.

Implementation results

Due to space constraints, please refer to the video for the implementation demo results; the implementation details will be explained during the presentation.
https://www.youtube.com/watch?v=zpdn_4kstuM