**Host Your Own AI Deep Research Agent with n8n,**

Apify and OpenAI o3

Workflow Overview

1. **Input Capture (Form Trigger):**
   * A user submits a research query through a form (using the "On form submission" node).
   * The form asks for a research query, along with parameters for how deep and broad the research should go.
2. **Variable Setup:**
   * The "Set Variables" node extracts key parameters (query, depth, breadth, and a unique request ID) from the form submission.
   * These values are used to control the recursive research process.
3. **Initial Query & SERP Generation:**
   * The "Set Initial Query" node prepares the initial search query based on the user's input and any prior conversation (if applicable).
   * The "Generate SERP Queries" node (using an LLM chain) takes the initial query and any previously learned insights to produce a list of search queries. The goal is to generate a diverse set of queries for exploring the research topic.
4. **Web Search:**
   * The workflow uses an HTTP Request node to call the Apify SERP scraper, which performs a Google search based on the generated queries.
   * The "Top 5 Organic Results" node filters the search results to keep only the top five organic entries.
5. **Content Extraction:**
   * A subsequent HTTP Request node ("Page Contents") scrapes the content of each search result page.
   * The content is then converted from HTML to Markdown, providing a cleaner text version for further processing.
6. **Learnings Extraction:**
   * The "DeepResearch Learnings" LLM chain processes the Markdown content from each search result to extract a set of key learnings.
   * This step is performed iteratively, controlled by the depth and breadth parameters.
   * All learnings are aggregated over iterations, and the final aggregated learnings are stored in the workflow’s state.
7. **Clarifying Questions:**
   * If needed, the workflow uses another LLM chain ("Clarifying Questions") to generate follow-up questions based on the initial query. This helps refine the research direction.
8. **Report Generation:**
   * Once the recursive research loop completes (determined by a depth limit), the final set of learnings is passed to another LLM chain ("DeepResearch Report").
   * The LLM generates a detailed research report in Markdown format that includes headings, lists, and tables, structured around the research query and accumulated learnings.
9. **Notion Integration:**
   * The generated report is then sent to a Notion page.
   * The "Create Row" node creates a new page in Notion with the report title and description.
   * The "Notion Block Generator" node converts parts of the report into Notion blocks (the JSON objects required by the Notion API), which are then uploaded to Notion via the "Upload to Notion Page" node.
10. **Final Steps & User Feedback:**
    * The workflow provides a final confirmation form ("Confirmation") that informs the user their report is being generated.
    * If no results are found during web scraping, an "Empty Response" node provides a fallback response.

Key Advanced Features

* **Recursive Looping:**  
  The workflow implements a recursive loop to generate and refine multiple search queries and accumulate learnings over several iterations. This is controlled by the user-defined depth and breadth parameters, allowing for scalable research.
* **Deep Learning Integration:**  
  Various LLM chains are used to:
  + Generate search queries.
  + Extract key learnings from web content.
  + Generate follow-up questions.
  + Finally, compile a detailed research report.
* **Notion API Integration:**  
  The workflow seamlessly integrates with Notion to store the final report, converting the report into the Notion block format required by the Notion API.
* **Error Handling & Fallbacks:**  
  The workflow includes multiple checks (e.g., "Has Results?", "Valid Pages") to ensure that data is only processed if it meets the necessary criteria, with fallback responses provided in case of errors.

Use Case

This template is ideal for researchers, analysts, and anyone looking to automate deep research tasks. By leveraging AI to generate, refine, and compile research findings from the web, it significantly reduces manual effort while delivering comprehensive, structured insights.

Customization

* **Prompt Customization:**  
  Users can customize the prompts used by the various LLM chains to better fit their research needs.
* **Search Depth and Breadth:**  
  The workflow allows for adjusting the depth (number of iterations) and breadth (number of search queries per iteration) to balance thoroughness against time and cost.
* **Integration with Other Tools:**  
  Although Notion is used for report storage in this template, the final output can be redirected to other platforms or databases as needed.